

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/EP04/014170

International filing date: 13 December 2004 (13.12.2004)

Document type: Certified copy of priority document

Document details: Country/Office: EP
Number: 04090127.4
Filing date: 01 April 2004 (01.04.2004)

Date of receipt at the International Bureau: 28 January 2005 (28.01.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse



Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

04090127.4

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk



Anmeldung Nr:
Application no.: 04090127.4
Demande no:

Anmeldetag:
Date of filing: 01.04.04
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Epigenomics AG
Kastanienallee 24
10435 Berlin
ALLEMAGNE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se référer à la description.)

PITX2 - a marker to predict survival of patients diagnosed with breast cell
proliferative disease

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)
revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12Q1/68

Am Anmeldetag benannte Vertragsstaaten/Contracting states designated at date of
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL
PL PT RO SE SI SK TR LI

Title

PITX2 - a marker to predict survival of patients diagnosed with breast cell proliferative disease

Abstract

The present invention relates to methods for predicting the survival of a human being diagnosed with a cell proliferative disorder of the breast tissues, characterized by a step of determining the expression level of PITX2 or the genetic or the epigenetic modifications of the genomic DNA associated with the gene PITX2.

The invention also relates to modified sequences, to oligonucleotides and/or PNA-oligomers which can be used within the described methods.

Field of the Invention**BREAST CANCER SURVIVAL**

In European and American women, breast cancer is the most frequently diagnosed cancer and the second leading cause of cancer death. In women aged 40-55, breast cancer is the leading cause of death (Greenlee *et al.*, 2000). In 2002, there were 204,000 new cases of breast cancer in the US and a comparable number in Europe.

Breast cancer is defined as the uncontrolled proliferation of cells within breast tissues. Breasts are comprised of 15 to 20 lobes joined together by ducts. Cancer arises most commonly in the duct, but is also found in the lobes with the rarest type of cancer termed inflammatory breast cancer. It will be appreciated by those skilled in the art that there exists a continuing need to improve methods of early detection, classification and treatment of breast cancers. In contrast to the detection of some other common cancers such as cervical and dermal there are inherent difficulties in classifying and detecting breast cancers.

The first step of any treatment is the assessment of the patient's condition comparative to defined classifications of the disease. However the value of such a system is inherently dependent upon the quality of the classification. Breast cancers are staged according to their

size, location and occurrence of metastasis. Methods of treatment include the use of surgery, radiation therapy, chemotherapy and endocrine therapy, which are also used as adjuvant therapies to surgery. In general, more aggressive disease should be treated with more aggressive therapies.

Although the vast majority of early cancers are operable, i.e. the tumor can be completely removed by surgery, about one third of the patients with lymph-node negative diseases and about 50-60% of patients with node-positive disease will develop metastases during follow-up.

Based on this observation, systemic adjuvant treatment has been introduced for both node-positive and node-negative breast cancers. Systemic adjuvant therapy is administered after surgical removal of the tumor, and has been shown to reduce the risk of recurrence significantly. Several types of adjuvant treatment are available: endocrine treatment, also called hormone treatment (for hormone receptor positive tumors), different chemotherapy regimens, and antibody treatments based on novel agents like Herceptin (an antibody to an epidermal growth factor receptor).

The growth of the majority of breast cancers (appr. 70-80%) is dependent on the presence of estrogen. Therefore, one important target for adjuvant therapy is the removal of estrogen (e.g. by ovarian ablation) or the blocking of its synthesis or the blocking of its actions on the tumor cells either by blocking the receptor with competing substances (e.g. Tamoxifen) or by inhibiting the conversion of androgen into estrogen (e.g. aromatase inhibitors). Endocrine treatment is thought to be efficient only in tumors that express hormone receptors (the estrogen receptor (ER) and/or the progesterone receptor (PR)). Currently, the vast majority of women with hormone receptor positive breast cancer receive some form of endocrine treatment, independent of their nodal status. The most frequently used drug is Tamoxifen.

However, even in hormone receptor positive patients, not all patients benefit from endocrine treatment. Adjuvant endocrine therapy reduces mortality rates by 22% while response rates to endocrine treatment in the advanced setting are 50 to 60%.

Since Tamoxifen has relatively few side effects, treatment may be justified even for patients with low likelihood of benefit. However, these patients may require additional, more aggressive adjuvant treatment. Even in earliest and least aggressive tumors, such as node-negative, hormone receptor positive tumours, about 21% of patients relapse within 10 years after initial diagnosis if they receive Tamoxifen monotherapy only, as adjuvant treatment (Lancet. 1998 May 16;351(9114):1451-67. Tamoxifen for early breast cancer: an overview of

the randomised trials. Early Breast Cancer Trialists' Collaborative Group.). Similarly, some patients with hormone receptor negative disease may be treated sufficiently with surgery and potentially radiotherapy alone, whereas others may require additional chemotherapy.

Several cytotoxic regimens have shown to be effective in reducing the risk of relapse in breast cancer (Mansour *et al.*, 1998). According to current treatment guidelines, most node-positive patients receive adjuvant chemotherapy both in the US and Europe, since the risk of relapse is considerable. Nevertheless, not all patients do relapse, and there is a proportion of patients who would never have relapsed even without chemotherapy, but who nevertheless receive chemotherapy due to the currently used criteria. In hormone receptor positive patients, chemotherapy is usually given before endocrine treatment, whereas hormone receptor negative patients receive only chemotherapy.

The situation for node-negative patients is particularly complex. In the US, cytotoxic chemotherapy is recommended for node-negative patients, if the tumor is larger than 1 cm. In Europe, chemotherapy is considered for the node-negative cases if one or more risk factors such as tumor size larger than 2 cm, negative hormone receptor status, or tumor grading of three or age <35 is present. In general, there is a tendency to select premenopausal women for additional chemotherapy whereas for postmenopausal women, chemotherapy is often omitted. Compared to endocrine treatment, in particular Tamoxifen or aromatase inhibitors, chemotherapy is highly toxic, with short-term side effects such as nausea, vomiting, bone marrow depression, and long-term effects such as cardiotoxicity and an increased risk for secondary cancers.

It is currently not clear which breast cancer patients should be selected for more aggressive therapy and which would do well without additional aggressive treatment, and clinicians agree that there is a large need for proper selection of patients. The difficulty of selecting the right patients for chemotherapy, and the lack of suitable criteria is also reflected by a recent study which showed that chemotherapy is used much less frequently than recommended, based on data from the New Mexico Tumor registry (Du *et al.*, 2003). This study provides substantial evidence that there is a need for better selection of patients for chemotherapy or other, more aggressive forms of breast cancer therapy.

This invention is about a new biomarker, which can be used to solve the problem described above. Based on the observation that methylation of the gene PITX2 (also known as PTX2) in breast tumor tissue, obtained from the surgically removed tumor, or obtained from biopsy material prior to the removal, is correlated with the survival time of breast cancer patients treated with Tamoxifen monotherapy, we invented a tool allowing a better selection of patients for more aggressive therapy, for example a cytotoxic therapy (chemotherapy) (besides or instead of an endocrine treatment like treatment with Tamoxifen or aromatase inhibitors) for breast cancer patients.

It can be concluded that the expression levels of the protein or mRNA also correlate with said prognosis. Therefore, the analysis of either the expression levels of PITX2 protein, or PITX2 mRNA or the analysis of the patient's individual genetic or epigenetic modification of the gene PITX2 – summarized as the analysis of expression of the gene PITX2 - may serve as a method for predicting the survival of a patient with breast cancer. Especially the invention relates to methods for predicting the survival of a patient with breast cancer who is treated with at least one adjuvant endocrine treatment, wherein endocrine treatment is meant to comprise any treatment targeting the estrogen receptor pathway or estrogen synthesis pathway or estrogen conversion pathway i.e., which is involved in estrogen metabolism, production or secretion.

In this context survival is meant to describe the time from diagnosis or start of treatment to an endpoint, which may be the time of death (considering any reason for death or only death from breast cancer), or the time of recurrence of breast cancer, which may be local or distant, or the time of occurrence of any breast cancer associated disease. Therefore “predicting the survival” is meant to comprise predicting the disease free survival, as well as the overall survival or any other consideration of time between diagnosis and endpoint of treatment.

PITX2 (also known as PTX2) is known to belong to the PTX subfamily of PTX1, PTX2, and PTX3 genes which define a novel family of transcription factors, within the paired-like class of homeodomain factors. The gene PITX2 (NM_000325) encodes the paired-like homeodomain transcription factor 2, which is known to be expressed during development of anterior structures such as the eye, teeth, and anterior pituitary.

Toyota *et al.*, (2001) (Blood 97: p 2823-9.) found hypermethylation of the PITX2 gene in a large proportion of acute myeloid leukemias. Furthermore, in this study hypermethylation of PITX2 is positively correlated to methylation of the ER gene.

Although the expression of PITX2 is associated with cell differentiation and proliferation it has no heretofore recognized role in carcinogenesis of breast cancer or responsiveness to endocrine treatment.

EXPRESSION ANALYSIS

The expression of a gene, or rather the protein encoded by the gene, can be studied on four different levels: firstly, protein expression levels can be determined directly, secondly, mRNA transcription levels can be determined, thirdly, epigenetic modifications, such as gene's DNA methylation profile or the gene's histone profile; can be analyzed, as methylation is often correlated with inhibited protein expression, and fourth, the gene itself may be analyzed for genetic modifications such as mutations, deletions, polymorphisms etc. influencing the expression of the gene product.

The levels of observation that have been studied by the methodological developments of recent years in molecular biology, are the genes themselves, the transcription of these genes into RNA, and the translation into the resulting proteins. However how the activation and inhibition of specific genes, in specific cells and tissues, at specific time points in the course of development of an individual are controlled, is correlatable to the degree and character of the methylation of the genes or respectively the genome. In this respect, pathogenic conditions may manifest themselves in a changed methylation pattern of individual genes or of the genome.

The four terms that apply to the fields of overall genome-wide analysis of all these biological processes are called: Proteomics, Transcriptomics, Epigenomics (or Methyloomics) and Genomics. Methods and techniques that can be used for studying expression or studying the modifications responsible for expression on all of these levels are well described in the literature and therefore known to a person skilled in the art. They are described in text books of molecular biology and in a large number of scientific journals.

How to analyze the protein expression of a single gene is prior art. It usually requires an antibody specific for the gene product of interest. Appropriate technologies would be ELISA or Immunohistochemistry.

The analysis of the level of mRNA also has been described sufficiently. These days the gold standard is the reverse transcriptase PCR.

To avoid duplication a more detailed description of the prior art relating to existing and well known technologies is given within the description of the invention, as it is part of the invention.

US patent application 2003/0198970 by Gareth Roberts lists some of the technologies and methods on how to determine a person's "genetic make up", i.e. the genetic modifications, such as deletions, polymorphisms, mutations etc. that may vary between individuals and describes the potential role of this genetic sequence information in the individual's variability in disease, response to therapy and prognosis. Epigenetic differences however are not mentioned. The gene PITX2 is listed within this application as one gene name out of a long and comprehensive list of about 2,500 other gene names, suggesting its expression could play a role in some kind of treatment response. However, this is simply an assumption based on speculation only, as no experiments are disclosed, which demonstrate any kind of relation between genetic modifications of PITX2 and an individual's variation in treatment response.

A less established area in this context is the field of epigenomics or epigenetics, i.e. the field concerned with analysis of DNA methylation patterns.

Methylation of DNA can play an important role in the control of gene expression in mammalian cells. DNA methyltransferases are involved in DNA methylation and catalyze the transfer of a methyl group from S-adenosylmethionine to cytosine residues to form 5-methylcytosine, a modified base that is found mostly at CpG sites in the genome. The presence of methylated CpG islands in the promoter region of genes can suppress their expression. This process may be due to the presence of 5-methylcytosine, which apparently interferes with the binding of transcription factors or other DNA-binding proteins to block transcription. In different types of tumors, aberrant or accidental methylation of CpG islands in the promoter region has been observed for many cancer-related genes, resulting in the silencing of their expression. Such genes include tumor suppressor genes, genes that suppress metastasis and angiogenesis, and genes that repair DNA (Momparker and Bovenzi (2000) J. Cell Physiol. 183:145-54).

In addition it has been described that DNA methylation may also play a role in the field of pharmacogenetics. A similar approach on how to apply information about genetic

modifications of the genome to the analysis of individual responses to treatment as was for example described by Gareth Roberts in US application 2003/0198970 was already subject of the application WO 02/037398, tailored to the application of information about *epigenetic* modifications of the genome, based on DNA methylation analysis, to guide treatment selection and to study individual's treatment responses.

An example for the applicability of this idea was given by Esteller et al. (Esteller et al. (2000) N Engl J Med. 2000 Nov 9;343(19):1350-4.), who demonstrated that methylation of the MGMT promoter in gliomas is a useful predictor of the responsiveness of the tumors to alkylating agents.

An example for the potential of analysis of epigenetic modifications, such as DNA methylation analysis, for the prediction of treatment response - related to breast cancer- was presented as a poster by Martens et al. at the San Antonio Breast Cancer Symposium, San Antonio, TX, December, 3-6, 2003. Breast cancer patients which have had their tumors removed by surgery and developed metastases at some point after the removal, were treated with Tamoxifen, an endocrine treatment drug. The primary tumor samples were analyzed for aberrant methylation patterns. The patients were then divided into two sub classes according to their objective tumor response: patients with progressive disease (which could be described as increasing metastasis size) and patients with complete or partial remission of the relapsed tumor (which could be described as decreasing metastasis size). It turned out, that those patients which had a tumor removed and experienced a remission (decrease in size) of the metastasis under endocrine treatment, had suffered from a tumor which showed a distinct pattern of DNA methylation at specific CpG sites, whereas patients which show progressive disease (did not experience a decrease but an increase in size of their metastases), under endocrine treatment, suffered from a tumor which did not show this distinct pattern of DNA methylation (but a different pattern) at these CpG sites. This is a clear indication, that the methylation pattern described in that study can serve as a **predictive** treatment response tool for an endocrine treatment, like tamoxifen. The results of this study, i.e. predictive biomarkers and assays therefore, are subject of patent application PCT/EP03/07827 [not yet published]: Method and nucleic acid for the analysis of breast cell proliferative disorders. Predictive markers as described above will also be called 'metastatic' markers in the context of this application.

Currently several predictive markers are under evaluation. As up to now most patients have received Tamoxifen as endocrine treatment most of the markers have been shown to be associated with response or resistance to tamoxifen. However, it is generally assumed that there is a large overlap between responders to one or the other endocrine treatment. In fact, ER and PR expression are used to select patients for any endocrine treatment. Among the markers which have been associated with TAM response is bcl-2. High bcl-2 levels showed promising correlation to TAM therapy response in patients with metastatic disease and prolonged survival and added valuable information to an ER negative patient subgroup (J Clin Oncology, 1997, 15 5: 1916-1922; Endocrine, 2000, 13(1):1-10). There is conflicting evidence regarding the independent predictive value of c-erbB2 (Her2/neu) overexpression in patients with advanced breast cancer that require further evaluation and verification (British J of Cancer, 1999, 79 (7/8):1220-1226; J Natl Cancer Inst, 1998, 90 (21): 1601-1608).

Other predictive markers include SRC-1 (steroid receptor coactivator-1), CGA gene over expression, cell kinetics and S phase fraction assays (Breast Cancer Res and Treat, 1998, 48:87-92; Oncogene, 2001, 20:6955-6959). Recently, uPA (Urokinase-type plasminogen activator) and PAI-1 (Plasminogen activator inhibitor type 1) together showed to be useful to define a subgroup of patients who have worse prognosis and who would benefit from adjuvant systemic therapy (J Clinical Oncology, 2002, 20 n° 4). However, all of these markers need further evaluations in prospective trials as none of them is yet a validated marker of response.

Also recently published was a study related to the **prognostic** power of methylation analysis in breast cancer patients. Müller et al. (Muller HM, Widschwendter A, Fiegl H, Ivarsson L, Goebel G, Perkmann E, Marth C, Widschwendter M. (2003) DNA methylation in serum of breast cancer patients: an independent prognostic marker. Cancer Res. 2003 Nov 15; 63(22): 7641-5.) reported about a set of genes, which can be used as biomarkers in patient pre-therapeutic sera for the prognosis of breast cancer. Specific aberrant methylation patterns of two genes found in DNA from pretreatment serum of cancer patients indicated whether their prognosis was good or bad. The DNA analyzed was not tumor DNA but serum DNA. Most likely the presence of a tumor-specific pattern indicates that tumor derived DNA is present, however, the absence of a specific methylation pattern, may be due to a tumor which does not show this methylation pattern, or a tumor which does not shed sufficient DNA into the blood

stream. Good or bad prognosis was defined as long or short "overall survival" after surgery, without adjuvant treatment. This result therefore relates to untreated patients, only.

These 'prognostic' markers are able to answer the question whether or not a breast cancer patient should get an aggressive adjuvant treatment like chemotherapy after removal of the tumor to avoid recurrence of cancer, i.e. occurrence of metastases.

However, none of these study results and none of these markers is able to answer the specific question raised above, whether or not a breast cancer patient should get adjuvant chemotherapy after removal of the tumor to avoid recurrence of cancer, i.e. occurrence of metastases in addition to endocrine treatment (with a drug like tamoxifen, or aromatase inhibitors).

A marker for a bad prognosis for cancer patients (without treatment), might not be applicable to a patient under adjuvant treatment with a drug like tamoxifen. Therefore the test would not be able to help deciding, whether chemotherapy, including all its side effects and inherent risks, is necessary or whether endocrine treatment is sufficient, because an endocrine treatment might change the prognosis from "bad" to "good".

The predictive 'metastatic' marker set described above, would be able to identify amongst all patient which relapse (develop metastases after surgery) those patients, which do not respond to endocrine treatment (by partial or complete remission of relapsed tumor). These markers however, cannot be applied to answer the question whether metastases will occur at all (after surgery of the primary tumor under endocrine treatment), and consequently whether it is advised to give adjuvant chemotherapy to avoid recurrence of cancer (i.e. relapse or occurrence of metastases).

In one aspect the present invention provides a marker, PITX2 (NM_000325), that can be used to answer that question and help guiding the decision whether or not an adjuvant chemotoxic therapy shall be subscribed in addition or instead of treatment with endocrines, like tamoxifen. A marker able to answer this question will also be called 'adjuvant' marker, in the context of this application.

In addition study results presented by Paik et al. at the San Antonio Breast Cancer Symposium, San Antonio, TX, December, 3-6, 2003 provide an answer to this question, by

analyzing the mRNA expression pattern of 16 genes with RT-PCR. They did not publish the identity of these 16 'adjuvant' markers.

For demonstration : The 'metastatic' test (use of a 'metastatic' marker) tells a patient that she will not respond to endocrine treatment when she develops metastases. But she does not know how high the likelihood is, that she will experience a relapse at all.

The 'prognostic' test (use of a 'prognostic' marker) tells a patient whether she will have a good or bad prognosis without any treatment. Even with a "bad prognosis" endocrine treatment might be enough.

The 'adjuvant test' (use of an 'adjuvant' marker) tells her whether she will or will not develop recurrence, without chemotherapy, even when treated with the standard -low side effected- endocrine treatment.

PITX2, however, which serves as an 'adjuvant marker' may also work as a 'prognostic marker', especially in hormone receptor negative women, which would not get any endocrine treatment at all.

5-methylcytosine is the most frequent covalent base modification in the DNA of eukaryotic cells. It plays a role, for example, in the regulation of the transcription, in genetic imprinting, and in tumorigenesis. Therefore, the identification of 5-methylcytosine as a component of genetic information is of considerable interest. However, 5-methylcytosine positions cannot be identified by sequencing since 5-methylcytosine has the same base pairing behaviour as cytosine. Moreover, the epigenetic information carried by 5-methylcytosine is completely lost during PCR amplification.

A relatively new and currently the most frequently used method for analysing DNA for 5-methylcytosine is based upon the specific reaction of bisulfite with cytosine which, upon subsequent alkaline hydrolysis, is converted to uracil which corresponds to thymidine in its base pairing behaviour. However, 5-methylcytosine remains unmodified under these conditions. Consequently, the original DNA is converted in such a manner that methylcytosine, which originally could not be distinguished from cytosine by its hybridisation behaviour, can now be detected as the only remaining cytosine using "normal" molecular biological techniques, for example, by amplification and hybridisation or sequencing. All of

these techniques are based on base pairing which can now be fully exploited. In terms of sensitivity, the prior art is defined by a method which encloses the DNA to be analysed in an agarose matrix, thus preventing the diffusion and renaturation of the DNA (bisulfite only reacts with single-stranded DNA), and which replaces all precipitation and purification steps with fast dialysis (Olek A, Oswald J, Walter J. A modified and improved method for bisulphite based cytosine methylation analysis. *Nucleic Acids Res.* 1996 Dec 15;24(24):5064-6). Using this method, it is possible to analyse individual cells, which illustrates the potential of the method. However, currently only individual regions of a length of up to approximately 3000 base pairs are analysed, a global analysis of cells for thousands of possible methylation events is not possible. However, this method cannot reliably analyse very small fragments from small sample quantities either. These are lost through the matrix in spite of the diffusion protection.

An overview of the further known methods of detecting 5-methylcytosine may be gathered from the following review article: Rein, T., DePamphilis, M. L., Zorbas, H., *Nucleic Acids Res.* 1998, 26, 2255.

To date, barring few exceptions (e.g., Zeschnigk M, Lich C, Buiting K, Doerfler W, Horsthemke B. A single-tube PCR test for the diagnosis of Angelman and Prader-Willi syndrome based on allelic methylation differences at the SNRPN locus. *Eur J Hum Genet.* 1997 Mar-Apr;5(2):94-8) the bisulfite technique is only used in research. Always, however, short, specific fragments of a known gene are amplified subsequent to a bisulfite treatment and either completely sequenced (Olek A, Walter J. The pre-implantation ontogeny of the H19 methylation imprint. *Nat Genet.* 1997 Nov;17(3):275-6) or individual cytosine positions are detected by a primer extension reaction (Gonzalzo ML, Jones PA. Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res.* 1997 Jun 15;25(12):2529-31, WO 95/00669) or by enzymatic digestion (Xiong Z, Laird PW. COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.* 1997 Jun 15;25(12):2532-4). In addition, detection by hybridisation has also been described (Olek et al., WO 99/28498).

Further publications dealing with the use of the bisulfite technique for methylation detection in individual genes are: Grigg G, Clark S. Sequencing 5-methylcytosine residues in genomic DNA. *Bioessays.* 1994 Jun;16(6):431-6, 431; Zeschnigk M, Schmitz B, Dittrich B, Buiting K,

Horsthemke B, Doerfler W. Imprinted segments in the human genome: different DNA methylation patterns in the Prader-Willi/Angelman syndrome region as determined by the genomic sequencing method. *Hum Mol Genet.* 1997 Mar;6(3):387-95; Feil R, Charlton J, Bird AP, Walter J, Reik W. Methylation analysis on individual chromosomes: improved protocol for bisulphite genomic sequencing. *Nucleic Acids Res.* 1994 Feb 25;22(4):695-6; Martin V, Ribieras S, Song-Wang X, Rio MC, Dante R. Genomic sequencing indicates a correlation between DNA hypomethylation in the 5' region of the pS2 gene and its expression in human breast cancer cell lines. *Gene.* 1995 May 19;157(1-2):261-4; WO 97/46705 and WO 95/15373.

An overview of the Prior Art in oligomer array manufacturing can be gathered from a special edition of *Nature Genetics* (*Nature Genetics Supplement*, Volume 21, January 1999), published in January 1999, and from the literature cited therein.

Fluorescently labelled probes are often used for the scanning of immobilised DNA arrays. The simple attachment of Cy3 and Cy5 dyes to the 5'-OH of the specific probe are particularly suitable for fluorescence labels. The detection of the fluorescence of the hybridised probes may be carried out, for example via a confocal microscope. Cy3 and Cy5 dyes, besides many others, are commercially available.

Matrix Assisted Laser Desorption Ionization Mass Spectrometry (MALDI-TOF) is a very efficient development for the analysis of biomolecules (Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem.* 1988 Oct 15;60(20):2299-301). An analyte is embedded in a light-absorbing matrix. The matrix is evaporated by a short laser pulse thus transporting the analyte molecule into the vapour phase in an unfragmented manner. The analyte is ionised by collisions with matrix molecules. An applied voltage accelerates the ions into a field-free flight tube. Due to their different masses, the ions are accelerated at different rates. Smaller ions reach the detector sooner than bigger ones.

MALDI-TOF spectrometry is excellently suited to the analysis of peptides and proteins. The analysis of nucleic acids is somewhat more difficult (Gut I G, Beck S. DNA and Matrix Assisted Laser Desorption Ionization Mass Spectrometry. *Current Innovations and Future Trends.* 1995, 1; 147-57). The sensitivity to nucleic acids is approximately 100 times worse

than to peptides and decreases disproportionally with increasing fragment size. For nucleic acids having a multiply negatively charged backbone, the ionisation process via the matrix is considerably less efficient. In MALDI-TOF spectrometry, the selection of the matrix plays an eminently important role. For the desorption of peptides, several very efficient matrixes have been found which produce a very fine crystallisation. There are now several responsive matrixes for DNA, however, the difference in sensitivity has not been reduced. The difference in sensitivity can be reduced by chemically modifying the DNA in such a manner that it becomes more similar to a peptide. Phosphorothioate nucleic acids in which the usual phosphates of the backbone are substituted with thiophosphates can be converted into a charge-neutral DNA using simple alkylation chemistry (Gut IG, Beck S. A procedure for selective DNA alkylation and detection by mass spectrometry. *Nucleic Acids Res.* 1995 Apr 25;23(8):1367-73). The coupling of a charge tag to this modified DNA results in an increase in sensitivity to the same level as that found for peptides. A further advantage of charge tagging is the increased stability of the analysis against impurities which make the detection of unmodified substrates considerably more difficult.

Genomic DNA is obtained from DNA of cell, tissue or other test samples using standard methods. This standard methodology is found in references such as Sambrook, Fritsch and Maniatis eds., *Molecular Cloning: A Laboratory Manual*, 1989.

DESCRIPTION

Characterisation of a breast cancer in terms of its predicted aggressiveness enables the physician to make an informed decision as to a therapeutic regimen with appropriate risk and benefit trade offs to the patient. Aggressiveness is taken to mean one or more of decreased patient survival or disease- or relapse-free survival, increased tumor-related complications and faster progression of tumor or metastases. According to the aggressiveness of the disease an appropriate treatment or treatments may be selected from the group consisting of chemotherapy, radiotherapy, surgery, biological therapy, immunotherapy, antibody treatments, treatments involving molecularly targeted drugs, estrogen receptor modulator treatments, estrogen receptor down-regulator treatments, aromatase inhibitors treatments, ovarian ablation, treatments providing LHRH analogues or other centrally acting drugs influencing estrogen production. Wherein a cancer is characterised as 'aggressive' it is particularly preferred that a treatment such as, but not limited to, chemotherapy is provided in addition to or instead of an endocrine targeting therapy.

Using the methods and nucleic acids described herein, statistically significant models of patient disease free or overall survival and/or disease progression can be developed and utilised to assist patients and clinicians in determining suitable treatment options to be included in the therapeutic regimen.

In one aspect the described method is to be used to assess the utility of therapeutic regimens comprising one or more treatments which is either an aggressive therapy such as chemotherapy or a treatment which targets the estrogen receptor pathway or is involved in estrogen metabolism, production or secretion as a therapy for patients suffering from a cell proliferative disorder of the breast tissues. In particular this aspect of the method enables the physician to determine which treatments may be used in addition to or instead of said endocrine treatment.

In a further aspect the described method enables the characterisation of the cell proliferative disorder in terms of aggressiveness, thereby enabling the physician to recommend suitable treatments. Thus, the present invention will be seen to reduce the problems associated with present breast cell proliferative disorder treatment response prediction methods.

Using the methods and nucleic acids as described herein, patient survival can be evaluated before or during treatment for a cell proliferative disorder of the breast tissues, in order to provide critical information to the patient and clinician as to the likely progression of the disease. It will be appreciated, therefore, that the methods and nucleic acids exemplified herein can serve to improve a patient's quality of life and odds of treatment success by allowing both patient and clinician a more accurate assessment of the patient's treatment options.

The method according to the definition may be used for the improved treatment of all breast cell proliferative disorder patients, both pre and post menopausal and independent of their node or estrogen receptor status. However, it is particularly preferred that said patients are node-negative and estrogen receptor positive.

The present invention makes available a method for the improved treatment and monitoring of breast cell proliferative disorders, by enabling the accurate prediction of a patient's survival without systemic therapy or with endocrine therapy comprising one or more treatments which target the estrogen receptor pathway or are involved in estrogen metabolism, production, or secretion.

In a particularly preferred embodiment, the method according to the invention enables the differentiation between patients who have a high risk of relapse under said endocrine therapy and those who have a low risk of relapse under said therapy.

The method particularly preferably enables the determination of a methylation pattern characteristic for a predicted survival time, in addition to the characterisation of tumors in terms of aggressiveness.

The method according to the invention may be used for the analysis of a wide variety of cell proliferative disorders of the breast tissues including, but not limited to, ductal carcinoma *in situ*, invasive ductal carcinoma, invasive lobular carcinoma, lobular carcinoma *in situ*, comedocarcinoma, inflammatory carcinoma, mucinous carcinoma, scirrhous carcinoma, colloid carcinoma, tubular carcinoma, medullary carcinoma, metaplastic carcinoma, and papillary carcinoma and papillary carcinoma *in situ*, undifferentiated or anaplastic carcinoma and Paget's disease of the breast.

The method according to the invention is particularly suited to the prediction of survival of breast cancer in the following treatment setting. In one embodiment, the method is applied to patients who receive endocrine pathway targeting treatment as secondary treatment to an initial non chemotherapeutical therapy, e.g. surgery (hereinafter referred to as the adjuvant setting) as illustrated in Figure 1. Such a treatment is often prescribed to patients suffering from Stage 1 to 3 breast carcinomas. In this embodiment patients survival times are predicted according to their gene expression or genetic or epigenetic modifications. By detecting patients with worse disease free survival times the physician may choose to recommend the patient for further treatment, instead of or in addition to the endocrine targeting therapy(s), in particular but not limited to, chemotherapy.

This invention is specifically about a new biomarker, PITX2, for patients diagnosed with breast cell proliferative disease, allowing the prediction of outcome without treatment, or with different therapies, like a cytotoxic therapy (chemotherapy) in addition or instead of (for example in hormone receptor negative patients) an endocrine treatment, like treatment with Tamoxifen or aromatase inhibitors, wherein the prediction is based on the patient's survival or clinical or pathological tumor response, or response measured with other surrogate parameters.

This invention therefore related to new methods or tools, for patients diagnosed with breast cell proliferative disease, allowing the evaluation of adjuvant therapy based on a prediction of outcome.

More specifically this invention provides new methods or tools, for patients diagnosed with breast cell proliferative disease, allowing the evaluation of adjuvant therapy, i.e. therapy after surgical removal of the tumor, like a cytotoxic therapy (chemotherapy) in addition to or instead of (for example in hormone receptor negative patients) an endocrine treatment, like treatment with Tamoxifen or aromatase inhibitors, wherein the evaluation is based on the prediction of the patient's survival.

One aspect of the invention is the provision of tools for predicting the survival of a patient diagnosed with a breast cell proliferative disease, such as breast cancer. These tools comprise methods for the analysis of either the expression levels of PITX2 protein, or PITX2 mRNA or the analysis of the patient's individual genetic or epigenetic modification of the gene PITX2 – summarized as the analysis of expression of the gene PITX2. Preferably the invention relates to methods for predicting the survival of a patient diagnosed with breast cancer. Preferably said patient is treated with at least one adjuvant endocrine treatment, wherein endocrine treatment is meant to comprise any treatment targeting the estrogen receptor pathway or estrogen synthesis pathway or estrogen conversion pathway i.e., which is involved in estrogen metabolism, production or secretion. Preferably the patient was treated with said adjuvant endocrine treatment after surgical removal of the tumor. Also preferably the survival is the disease free survival.

Especially preferred are methods applied for the prediction of the disease free survival of a patient diagnosed with breast cancer under adjuvant endocrine treatment after surgical tumor removal. Even more preferred are those methods, which analyze the DNA methylation profile of the genomic region associated with the gene PITX2. Especially preferred is the analysis of the DNA methylation profile of the genomic sequence given in SEQ ID 1.

This methodology presents further improvements over the state of the art in that the method may be applied to any subject, independent of the estrogen and/or progesterone receptor status. Therefore in a preferred embodiment, the subject is not required to have been tested for estrogen or progesterone receptor status.

The object of the invention is achieved by means of the analysis of the methylation pattern of PITX2 and/or its regulatory region. In a particularly preferred embodiment the sequence of said gene comprises SEQ ID 1 and the sequence complementary thereto.

In one embodiment the object of the invention is the prediction of survival under a treatment which targets the estrogen receptor pathway or is involved in estrogen metabolism, production or secretion. This is achieved by analysis of the expression pattern of PITX2 and wherein it is further preferred that the sequence of said gene comprises SEQ ID NO: 1 .

In one aspect the invention discloses novel methods utilizing the gene PITX2 for the prediction of survival of a patient diagnosed with a breast cell proliferative disease. In a preferred embodiment said patient diagnosed with a breast cell proliferative disease is treated with adjuvant endocrine monotherapy.

The invention discloses the use of the gene PITX2, as well as its promoter and regulatory elements as a prognostic marker for survival of breast cancer patients. It is preferred that these patients are treated with adjuvant endocrine monotherapy. More specifically, the disclosed matter shows the applicability of said gene to answer the question and help guiding the decision whether or not an adjuvant chemotoxic therapy shall be subscribed, preferably in addition of endocrine treatment, like the treatment with tamoxifen or aromatase inhibitors.

In one aspect of the invention, the disclosed matter provides novel nucleic acid sequences useful for the analysis of methylation within said gene, other aspects provide novel uses of the gene and the gene product as well as methods, assays and kits directed to prognosing the survival of a patient diagnosed with breast cell proliferative disease. Preferably a patient which is treated with adjuvant endocrine monotherapy.

In one embodiment the method discloses the use of the gene PITX2 as a marker for the prognosis of the survival of a patient suffering from a breast cell proliferative disease. Preferably said patient is treated with adjuvant endocrine monotherapy. Said use of the gene may be enabled by means of any analysis of the expression of the gene, by means of mRNA expression analysis or protein expression analysis or by analysis of its genetic modifications leading to an altered expression.

However, in the most preferred embodiment of the invention, prediction of the survival of a patient diagnosed with breast cell proliferative disease, preferably treated with adjuvant endocrine monotherapy, is enabled by means of analysis of the methylation status of CpG sites within the gene PITX2 and its promoter or regulatory elements.

To detect the presence of mRNA encoding PITX2 in a detection system for breast cancer relapse, a sample is obtained from a patient. The sample can be a tumor tissue sample from the surgically removed tumor, a biopsy sample or a sample of blood, plasma, serum or the like. The sample may be treated to extract the nucleic acids contained therein. The resulting nucleic acid from the sample is subjected to gel electrophoresis or other separation techniques. Detection involves contacting the nucleic acids and in particular the mRNA of the sample with a DNA sequence serving as a probe to form hybrid duplexes. The stringency of hybridisation is determined by a number of factors during hybridisation and during the washing procedure, including temperature, ionic strength, length of time and concentration of formamide. These factors are outlined in, for example, Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, 2nd ed., 1989). Detection of the resulting duplex is usually accomplished by the use of labelled probes. Alternatively, the probe may be unlabeled, but may be detectable by specific binding with a ligand which is labelled, either directly or indirectly. Suitable labels and methods for labelling probes and ligands are known in the art, and include, for example, radioactive labels which may be incorporated by known methods (e.g., nick translation or kinasing), biotin, fluorescent groups, chemiluminescent groups (e.g., dioxetanes, particularly triggered dioxetanes), enzymes, antibodies, and the like.

In order to increase the sensitivity of the detection in a sample of mRNA encoding PITX2, the technique of reverse transcription/polymerisation chain reaction can be used to amplify cDNA transcribed from mRNA encoding PITX2. The method of reverse transcription /PCR is well known in the art (for example, see Watson and Fleming, *supra*).

The reverse transcription /PCR method can be performed as follows. Total cellular RNA is isolated by, for example, the standard guanidium isothiocyanate method and the total RNA is reverse transcribed. The reverse transcription method involves synthesis of DNA on a template of RNA using a reverse transcriptase enzyme and a 3' end primer. Typically, the primer contains an oligo(dT) sequence. The cDNA thus produced is then amplified using the PCR method and PITX2 specific primers. (Belyavsky et al, *Nucl Acid Res* 17:2919-2932,

1989; Krug and Berger, Methods in Enzymology, Academic Press, N.Y., Vol.152, pp. 316-325, 1987 which are incorporated by reference)

The present invention may also be described in certain embodiments as a kit for use in prognosing the survival of a breast cancer patient before or after surgical tumor removal with or without adjuvant endocrine monotherapy state through testing of a biological sample. A representative kit may comprise one or more nucleic acid segments as described above that selectively hybridise to PITX2 mRNA and a container for each of the one or more nucleic acid segments. In certain embodiments the nucleic acid segments may be combined in a single tube. In further embodiments, the nucleic acid segments may also include a pair of primers for amplifying the target mRNA. Such kits may also include any buffers, solutions, solvents, enzymes, nucleotides, or other components for hybridisation, amplification or detection reactions. Preferred kit components include reagents for reverse transcription-PCR, in situ hybridisation, Northern analysis and/or RPA.

The present invention further provides for methods to detect the presence of the polypeptide, PITX2, in a sample obtained from a patient. Any method known in the art for detecting proteins can be used. Such methods include, but are not limited to immunodiffusion, immunoelectrophoresis, immunochemical methods, binder-ligand assays, immunohistochemical techniques, agglutination and complement assays. (for example see Basic and Clinical Immunology, Sites and Terr, eds., Appleton & Lange, Norwalk, Conn. pp 217-262, 1991 which is incorporated by reference). Preferred are binder-ligand immunoassay methods including reacting antibodies with an epitope or epitopes of PITX2 and competitively displacing a labelled PITX2 protein or derivative thereof.

Certain embodiments of the present invention comprise the use of antibodies specific to the polypeptide encoded by the PITX2 gene. Such antibodies may be useful for prognosing the survival of a breast cancer patient preferably under adjuvant endocrine monotherapy by comparing a patient's levels of PITX2 marker expression to expression of the same marker in normal individuals. In certain embodiments production of monoclonal or polyclonal antibodies can be induced by the use of the PITX2 polypeptide as antigen. Such antibodies may in turn be used to detect expressed proteins as markers for prognosis of relapse of a breast cancer patient under adjuvant endocrine monotherapy. The levels of such proteins present in the peripheral blood of a patient may be quantified by conventional methods.

Antibody-protein binding may be detected and quantified by a variety of means known in the art, such as labelling with fluorescent or radioactive ligands. The invention further comprises kits for performing the above-mentioned procedures, wherein such kits contain antibodies specific for the PITX2 polypeptides.

Numerous competitive and non-competitive protein binding immunoassays are well known in the art. Antibodies employed in such assays may be unlabeled, for example as used in agglutination tests, or labelled for use a wide variety of assay methods. Labels that can be used include radionuclides, enzymes, fluorescers, chemilumescers, enzyme substrates or co-factors, enzyme inhibitors, particles, dyes and the like for use in radioimmunoassay (RIA), enzyme immunoassays, e.g., enzyme-linked immunosorbent assay (ELISA), fluorescent immunoassays and the like. Polyclonal or monoclonal antibodies to PITX2 or an epitope thereof can be made for use in immunoassays by any of a number of methods known in the art. One approach for preparing antibodies to a protein is the selection and preparation of an amino acid sequence of all or part of the protein, chemically synthesising the sequence and injecting it into an appropriate animal, usually a rabbit or a mouse (Milstein and Kohler Nature 256:495-497, 1975; Gutfre and Milstein, Methods in Enzymology: Immunochemical Techniques 73:1-46, Langone and Banatis eds., Academic Press, 1981 which are incorporated by reference). Methods for preparation of PITX2 or an epitope thereof include, but are not limited to chemical synthesis, recombinant DNA techniques or isolation from biological samples.

The invention provides significant improvements over the state of the art in that there are currently no markers known to the public which can be used to detect the likelihood of relapse or of survival of a breast cancer patient under adjuvant endocrine monotherapy, neither from tissue samples nor from body fluid samples.

Also, no methylation marker is known which can be used to detect the likelihood of relapse or of survival of a breast cancer patient. Especially, no methylation marker is known which can be used to detect the likelihood of relapse or of survival of a breast cancer patient under adjuvant endocrine monotherapy, neither from tissue samples nor from body fluid samples.

The objective of the invention can also be achieved by analysis of the methylation state of the CpG dinucleotides within the genomic sequence according to SEQ ID NO: 1 and sequences complementary thereto. SEQ ID NO: 1 discloses the gene PITX2 and its promoter and

regulatory elements, wherein said fragment comprises CpG dinucleotides exhibiting a disease specific methylation pattern. The methylation pattern of the gene PITX2 and its promoter and regulatory elements have heretofore not been analysed with regard to prognosis of survival of a patient diagnosed with a breast cell proliferative disorder . Due to the degeneracy of the genetic code, the sequence as identified in SEQ ID NO: 1 should be interpreted so as to include all substantially similar and equivalent sequences upstream of the promoter region of a gene which encodes a polypeptide with the biological activity of that encoded by PITX2.

In a preferred embodiment of the method, the objective of the invention is achieved by analysis of a nucleic acid comprising a sequence of at least 18 bases in length according to one of SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto.

The sequences of SEQ ID NOS: 2 to 5 provide modified versions of the nucleic acid according to SEQ ID NO: 1, wherein the conversion of said sequence results in the synthesis of a nucleic acid having a sequence that is unique and distinct from SEQ ID NO: 1 as follows. (*see also* the following TABLE 1): SEQ ID NO: 1, sense DNA strand of PITX2 gene and its promoter and regulatory elements; SEQ ID NO: 2, converted SEQ ID NO: 1, wherein "C" converted to "T," but "CpG" remains "CpG." (*i.e.*, corresponds to case where, for SEQ ID NO: 1, all "C" residues of CpG dinucleotide sequences are methylated and are thus not converted); SEQ ID NO: 3, complement of SEQ ID NO: 1, wherein "C" converted to "T," but "CpG" remains "CpG" (*i.e.*, corresponds to case where, for the complement (*antisense* strand) of SEQ ID NO: 1, all "C" residues of CpG dinucleotide sequences are methylated and are thus not converted); SEQ ID NO: 4, converted SEQ ID NO: 1, wherein "C" converted to "T" for all "C" residues, including those of "CpG" dinucleotide sequences (*i.e.*, corresponds to case where, for SEQ ID NO: 1, all "C" residues of CG dinucleotide sequences are unmethylated); SEQ ID NO: 5, complement of SEQ ID NO: 1, wherein "C" converted to "T" for all "C" residues, including those of "CpG" dinucleotide sequences (*i.e.*, corresponds to case where, for the complement (*antisense* strand) of SEQ ID NO: 1, all "C" residues of CpG dinucleotide sequences are unmethylated).

TABLE 1. Description of SEQ ID NOS: 1 to 5

SEQ ID NO	Relationship to SEQ ID NO:1	Nature of cytosine base conversion
SEQ ID NO:1	Sense strand (PITX2 gene including promoter and regulatory elements)	None; untreated sequence
SEQ ID NO:2	Converted sense strand	"C" to "T," but "CpG" remains "CpG" (all "C" residues of CpGs are methylated)
SEQ ID NO:3	Converted antisense strand	"C" to "T," but "CpG" remains "CpG" (all "C" residues of CpGs are methylated)
SEQ ID NO:4	Converted sense strand	"C" to "T" for all "C" residues (all "C" residues of CpGs are <u>un</u> methylated)
SEQ ID NO:5	Converted antisense strand	"C" to "T" for all "C" residues (all "C" residues of CpGs are <u>un</u> methylated)

Significantly, heretofore, the nucleic acid sequences and molecules according to SEQ ID NO: 1 to SEQ ID NO: 5 were not implicated in or connected with the ascertainment of the prognosis of breast cancer relapse or survival.

The described invention further discloses oligonucleotides or oligomers for detecting the cytosine methylation state within pretreated DNA, according to SEQ ID NO: 2 to SEQ ID NO: 5. Said oligonucleotides or oligomers comprising a nucleic acid sequence having a length of at least nine (9) nucleotides which hybridise, under moderately stringent or stringent conditions (as defined herein above), to a pretreated nucleic acid sequence according to SEQ ID NO: 2 to SEQ ID NO: 5 and/or sequences complementary thereto.

Thus, the present invention includes nucleic acid molecules (*e.g.*, oligonucleotides and peptide nucleic acid (PNA) molecules (PNA-oligomers)) that hybridise under moderately stringent and/or stringent hybridisation conditions to all or a portion of the sequences of SEQ ID NOS: 2 to 5, or to the complements thereof. The hybridising portion of the hybridising nucleic acids is typically at least 9, 15, 20, 25, 30 or 35 nucleotides in length. However, longer molecules have inventive utility, and are thus within the scope of the present invention.

Preferably, the hybridising portion of the inventive hybridising nucleic acids is at least 95%, or at least 98%, or 100% identical to the sequence, or to a portion thereof of SEQ ID NOS: 2 to 5, or to the complements thereof.

Hybridising nucleic acids of the type described herein can be used, for example, as a primer (*e.g.*, a PCR primer), or a diagnostic and/or prognostic probe or primer. Preferably, hybridisation of the oligonucleotide probe to a nucleic acid sample is performed under stringent conditions and the probe is 100% identical to the target sequence. Nucleic acid duplex or hybrid stability is expressed as the melting temperature or T_m , which is the temperature at which a probe dissociates from a target DNA. This melting temperature is used to define the required stringency conditions.

For target sequences that are related and substantially identical to the corresponding sequence of SEQ ID NO: 1 (such as PITX2 allelic variants and SNPs), rather than identical, it is useful to first establish the lowest temperature at which only homologous hybridisation occurs with a particular concentration of salt (*e.g.*, SSC or SSPE). Then, assuming that 1% mismatching results in a 1°C decrease in the T_m , the temperature of the final wash in the hybridisation reaction is reduced accordingly (for example, if sequences having > 95% identity with the probe are sought, the final wash temperature is decreased by 5°C). In practice, the change in T_m can be between 0.5°C and 1.5°C per 1% mismatch.

Examples of inventive oligonucleotides of length X (in nucleotides), as indicated by polynucleotide positions with reference to, *e.g.*, SEQ ID NO: 1, include those corresponding to sets of consecutively overlapping oligonucleotides of length X, where the oligonucleotides within each consecutively overlapping set (corresponding to a given X value) are defined as the finite set of Z oligonucleotides from nucleotide positions:

n to (n + (X-1));

where n=1, 2, 3,...(Y-(X-1));

where Y equals the length (nucleotides or base pairs) of SEQ ID NO: 1 ;

where X equals the common length (in nucleotides) of each oligonucleotide in the set (*e.g.*, X=20 for a set of consecutively overlapping 20-mers); and

where the number (Z) of consecutively overlapping oligomers of length X for a given SEQ ID NO of length Y is equal to Y-(X-1). For example $Z=2,785-19=2,766$ for either sense or antisense sets of SEQ ID NO: 1, where X=20.

Preferably, the set is limited to those oligomers that comprise at least one CpG, Cpa or tpG dinucleotide, wherein 'Cpa' is indicating that said Cpa hybridises to a position (tpG) which was a CpG prior to bisulfite conversion and is a TpG now; and wherein 'tpG' is indicating that said tpG hybridises to a position (Cpa) which is the complementary to a position (tpG) which was a CpG prior to bisulfite conversion and is a TpG now.

The present invention encompasses, for *each* of SEQ ID NOS: 2 to 5 (sense and antisense), multiple consecutively overlapping sets of oligonucleotides or modified oligonucleotides of length X, where, *e.g.*, X= 9, 10, 17, 20, 22, 23, 25, 27, 30 or 35 nucleotides.

The oligonucleotides or oligomers according to the present invention constitute effective tools useful to ascertain genetic and epigenetic parameters of the genomic sequence corresponding to SEQ ID NO: 1. Preferred sets of such oligonucleotides or modified oligonucleotides of length X are those consecutively overlapping sets of oligomers corresponding to SEQ ID NOS:1-5 (and to the complements thereof). Preferably, said oligomers comprise at least one CpG, tpG or Cpa dinucleotide.

Particularly preferred oligonucleotides or oligomers according to the present invention are those in which the cytosine of the CpG dinucleotide (or of the corresponding converted TpG or CpA dinucleotide) sequences is within the middle third of the oligonucleotide; that is, where the oligonucleotide is, for example, 13 bases in length, the CpG, TpG or CpA dinucleotide is positioned within the fifth to ninth nucleotide from the 5'-end.

The oligonucleotides of the invention can also be modified by chemically linking the oligonucleotide to one or more moieties or conjugates to enhance the activity, stability or detection of the oligonucleotide. Such moieties or conjugates include chromophores, fluorophores, lipids such as cholesterol, cholic acid, thioether, aliphatic chains, phospholipids, polyamines, polyethylene glycol (PEG), palmityl moieties, and others as disclosed in, for example, United States Patent Numbers 5,514,758, 5,565,552, 5,567,810, 5,574,142, 5,585,481, 5,587,371, 5,597,696 and 5,958,773. The probes may also exist in the form of a PNA (peptide nucleic acid) which has particularly preferred pairing properties. Thus, the oligonucleotide may include other appended groups such as peptides, and may include hybridisation-triggered cleavage agents (Krol et al., *BioTechniques* 6:958-976, 1988) or

intercalating agents (Zon, *Pharm. Res.* 5:539-549, 1988). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a chromophore, fluorophor, peptide, hybridisation-triggered cross-linking agent, transport agent, hybridisation-triggered cleavage agent, etc.

The oligonucleotide may also comprise at least one art-recognized modified sugar and/or base moiety, or may comprise a modified backbone or non-natural internucleoside linkage.

The oligomers according to the present invention are normally used in so called "sets" which contain at least one oligomer for analysis of each of the CpG dinucleotides of a genomic sequence comprising SEQ ID NO: 1 and sequences complementary thereto or to their corresponding CG, tG or Ca dinucleotide within the pretreated nucleic acids according to SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto, wherein a 't' indicates a nucleotide which converted from a cytosine into a thymine and wherein 'a' indicates the complementary nucleotide to such a converted thymine. Preferred is a set which contains at least one oligomer for each of the CpG dinucleotides within the gene PITX2 and it's promoter and regulatory elements in both the pretreated and genomic versions of said gene, SEQ ID NO: 2 to 5 and SEQ ID NO: 1, respectively. However, it is anticipated that for economic or other factors it may be preferable to analyse a limited selection of the CpG dinucleotides within said sequences and the contents of the set of oligonucleotides should be altered accordingly. Therefore, the present invention moreover relates to a set of at least 3 n (oligonucleotides and/or PNA-oligomers) used for detecting the cytosine methylation state in pretreated genomic DNA (SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto) and genomic DNA (SEQ ID NO: 1 and sequences complementary thereto). These probes enable diagnosis and/or therapy of genetic and epigenetic parameters of cell proliferative disorders. The set of oligomers may also be used for detecting single nucleotide polymorphisms (SNPs) in pretreated genomic DNA (SEQ ID NO: 2 to SEQ ID NO: 5, and sequences complementary thereto) and genomic DNA (SEQ ID NO: 1, and sequences complementary thereto).

Moreover, the present invention makes available a set of at least two oligonucleotides which can be used as so-called "primer oligonucleotides" for amplifying DNA sequences of one of SEQ ID NO: 1 to SEQ ID NO: 5 and sequences complementary thereto, or segments thereof.

In the case of the sets of oligonucleotides according to the present invention, it is preferred that at least one and more preferably all members of the set of oligonucleotides is bound to a solid phase.

According to the present invention, it is preferred that an arrangement of different oligonucleotides and/or PNA-oligomers (a so-called "array") made available by the present invention is present in a manner that it is likewise bound to a solid phase. This array of different oligonucleotide- and/or PNA-oligomer sequences can be characterised in that it is arranged on the solid phase in the form of a rectangular or hexagonal lattice. The solid phase surface is preferably composed of silicon, glass, polystyrene, aluminium, steel, iron, copper, nickel, silver, or gold. However, nitrocellulose as well as plastics such as nylon which can exist in the form of pellets or also as resin matrices may also be used.

Therefore, a further subject matter of the present invention is a method for manufacturing an array fixed to a carrier material for analysis in connection with cell proliferative disorders, in which method at least one oligomer according to the present invention is coupled to a solid phase. Methods for manufacturing such arrays are known, for example, from US Patent 5,744,305 by means of solid-phase chemistry and photolabile protecting groups.

A further subject matter of the present invention relates to a DNA chip for the analysis of cell proliferative disorders. DNA chips are known, for example, in US Patent 5,837,832.

The described invention further provides a composition of matter useful for prognosing the relapse of breast cancer patients. Said composition comprising at least one nucleic acid 18 base pairs in length of a segment of the nucleic acid sequence disclosed in SEQ ID NO: 2 to 5, and one or more substances taken from the group comprising :

1-5 mM Magnesium Chloride, 100-500 μ M dNTP, 0.5-5 units of taq polymerase, bovine serum albumen, an oligomer in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridises under moderately stringent or stringent conditions to a pretreated genomic DNA according to one of the SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto. It is preferred that said composition of matter comprises a buffer solution appropriate for the stabilisation of said nucleic acid in an

aqueous solution and enabling polymerase based reactions within said solution.. Suitable buffers are known in the art and commercially available.

The present invention further provides a method for conducting an assay in order to ascertain genetic and/or epigenetic parameters of the gene PITX2 and its promoter and regulatory elements. Most preferably the assay according to the following method is used in order to detect methylation within the gene PITX2 wherein said methylated nucleic acids are present in a solution further comprising an excess of background DNA, wherein the background DNA is present in between 100 to 1000 times the concentration of the DNA to be detected. Said method comprising contacting a nucleic acid sample obtained from said subject with at least one reagent or a series of reagents, wherein said reagent or series of reagents, distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid.

Preferably, said method comprises the following steps: In the first *step*, a sample of the tissue to be analysed is obtained. The source may be any suitable source, preferably, the source of the sample is selected from the group consisting of histological slides, biopsies, paraffin-embedded tissue, bodily fluids, plasma, serum, stool, urine, blood, nipple aspirate and combinations thereof. Preferably, the source is tumor tissue, biopsies, serum, urine, blood or nipple aspirate. The most preferred source, is the tumor sample, surgically removed from the patient or a biopsy sample of said patient.

The DNA is then isolated from the sample. Extraction may be by means that are standard to one skilled in the art, including the use of detergent lysates, sonification and vortexing with glass beads. Once the nucleic acids have been extracted, the genomic double stranded DNA is used in the analysis.

In the second step of the method, the genomic DNA sample is treated in such a manner that cytosine bases which are unmethylated at the 5'-position are converted to uracil, thymine, or another base which is dissimilar to cytosine in terms of hybridisation behaviour. This will be understood as 'pretreatment' herein.

The above described treatment of genomic DNA is preferably carried out with bisulfite (hydrogen sulfite, disulfite) and subsequent alkaline hydrolysis which results in a conversion of non-methylated cytosine nucleobases to uracil or to another base which is dissimilar to

cytosine in terms of base pairing behaviour. Enclosing the DNA to be analysed in an agarose matrix, thereby preventing the diffusion and renaturation of the DNA (bisulfite only reacts with single-stranded DNA), and replacing all precipitation and purification steps with fast dialysis (Olek A, et al., A modified and improved method for bisulfite based cytosine methylation analysis, *Nucleic Acids Res.* 24:5064-6, 1996). It is further preferred that the bisulfite treatment is carried out in the presence of a radical scavenger or DNA denaturing agent.

In the third step of the method, fragments of the pretreated DNA are amplified. Wherein the source of the DNA is free DNA from serum, or DNA extracted from paraffin it is particularly preferred that the size of the amplificate fragment is between 100 and 200 base pairs in length, and wherein said DNA source is extracted from cellular sources (e.g. tissues, biopsies, cell lines) it is preferred that the amplificate is between 100 and 350 base pairs in length. It is particularly preferred that said amplificates comprise at least one 20 base pair sequence comprising at least three CpG dinucleotides. Said amplification is carried out using sets of primer oligonucleotides according to the present invention, and a preferably heat-stable polymerase. The amplification of several DNA segments can be carried out simultaneously in one and the same reaction vessel, in one embodiment of the method preferably six or more fragments are amplified simultaneously. Typically, the amplification is carried out using a polymerase chain reaction (PCR). The set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridise under stringent or highly stringent conditions to an at least 18-base-pair long segment of the base sequences of SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto.

In an alternate embodiment of the method, the methylation status of preselected CpG positions within the nucleic acid sequences comprising SEQ ID NO: 2 to SEQ ID NO: 5 may be detected by use of methylation-specific primer oligonucleotides. This technique (MSP) has been described in United States Patent No. 6,265,171 to Herman. The use of methylation status specific primers for the amplification of bisulfite treated DNA allows the differentiation between methylated and unmethylated nucleic acids. MSP primers pairs contain at least one primer which hybridises to a bisulfite treated CpG dinucleotide. Therefore, the sequence of said primers comprises at least one CpG, TpG or CpA dinucleotide. MSP primers specific for non-methylated DNA contain a "T" at the 3' position of the C position in the CpG. Preferably, therefore, the base sequence of said primers is required to comprise a sequence having a

length of at least 18 nucleotides which hybridises to a pretreated nucleic acid sequence according to SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG, tpG or Cpa dinucleotide. In this embodiment of the method according to the invention it is particularly preferred that the MSP primers comprise between 2 and 4 CpG , tpG or Cpa dinucleotides. It is further preferred that said dinucleotides are located within the 3' half of the primer e.g. wherein a primer is 18 bases in length the specified dinucleotides are located within the first 9 bases from the 3' end of the molecule. In addition to the CpG , tpG or Cpa dinucleotides it is further preferred that said primers should further comprise several bisulfite converted bases (i.e. cytosine converted to thymine, or on the hybridising strand, guanine converted to adenosine). In a further preferred embodiment said primers are designed so as to comprise no more than 2 cytosine or guanine bases.

In one embodiment of the method the primers may be selected from the group consisting of SEQ ID NO: 6 to SEQ ID NO: 10.

The fragments obtained by means of the amplification can carry a directly or indirectly detectable label. Preferred are labels in the form of fluorescence labels, radionuclides, or detachable molecule fragments having a typical mass which can be detected in a mass spectrometer. Where said labels are mass labels, it is preferred that the labelled amplicates have a single positive or negative net charge, allowing for better detectability in the mass spectrometer. The detection may be carried out and visualised by means of, e.g., matrix assisted laser desorption/ionisation mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI).

Matrix Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-TOF) is a very efficient development for the analysis of biomolecules (Karas & Hillenkamp, *Anal Chem.*, 60:2299-301, 1988). An analyte is embedded in a light-absorbing matrix. The matrix is evaporated by a short laser pulse thus transporting the analyte molecule into the vapour phase in an unfragmented manner. The analyte is ionised by collisions with matrix molecules. An applied voltage accelerates the ions into a field-free flight tube. Due to their different masses, the ions are accelerated at different rates. Smaller ions reach the detector sooner than bigger ones. MALDI-TOF spectrometry is well suited to the analysis of peptides and proteins. The analysis of nucleic acids is somewhat more difficult (Gut & Beck, *Current Innovations and*

Future Trends, 1:147-57, 1995). The sensitivity with respect to nucleic acid analysis is approximately 100-times less than for peptides, and decreases disproportionately with increasing fragment size. Moreover, for nucleic acids having a multiply negatively charged backbone, the ionisation process via the matrix is considerably less efficient. In MALDI-TOF spectrometry, the selection of the matrix plays an eminently important role. For the desorption of peptides, several very efficient matrixes have been found which produce a very fine crystallisation. There are now several responsive matrixes for DNA, however, the difference in sensitivity between peptides and nucleic acids has not been reduced. This difference in sensitivity can be reduced, however, by chemically modifying the DNA in such a manner that it becomes more similar to a peptide. For example, phosphorothioate nucleic acids, in which the usual phosphates of the backbone are substituted with thiophosphates, can be converted into a charge-neutral DNA using simple alkylation chemistry (Gut & Beck, *Nucleic Acids Res.* 23: 1367-73, 1995). The coupling of a charge tag to this modified DNA results in an increase in MALDI-TOF sensitivity to the same level as that found for peptides. A further advantage of charge tagging is the increased stability of the analysis against impurities, which makes the detection of unmodified substrates considerably more difficult.

In a particularly preferred embodiment of the method the amplification of step three is carried out in the presence of at least one species of blocker oligonucleotides. The use of such blocker oligonucleotides has been described by Yu et al., *BioTechniques* 23:714-720, 1997. The use of blocking oligonucleotides enables the improved specificity of the amplification of a subpopulation of nucleic acids. Blocking probes hybridised to a nucleic acid suppress, or hinder the polymerase mediated amplification of said nucleic acid. In one embodiment of the method blocking oligonucleotides are designed so as to hybridise to background DNA. In a further embodiment of the method said oligonucleotides are designed so as to hinder or suppress the amplification of unmethylated nucleic acids as opposed to methylated nucleic acids or vice versa.

Blocking probe oligonucleotides are hybridised to the bisulfite treated nucleic acid concurrently with the PCR primers. PCR amplification of the nucleic acid is terminated at the 5' position of the blocking probe, such that amplification of a nucleic acid is suppressed where the complementary sequence to the blocking probe is present. The probes may be designed to hybridise to the bisulfite treated nucleic acid in a methylation status specific manner. For example, for detection of methylated nucleic acids within a population of unmethylated

nucleic acids, suppression of the amplification of nucleic acids which are unmethylated at the position in question would be carried out by the use of blocking probes comprising a 'TpG' at the position in question, as opposed to a 'CpG.' In one embodiment of the method the sequence of said blocking oligonucleotides should be identical or complementary to molecule is complementary or identical to a sequence at least 18 base pairs in length selected from the group consisting of SEQ ID NOS: 2 to 5, preferably comprising one or more CpG, TpG or CpA dinucleotides . In one embodiment of the method the sequence of said oligonucleotides is selected from the group consisting SEQ ID NO: 15 and SEQ ID NO: 16 and sequences complementary thereto.

For PCR methods using blocker oligonucleotides, efficient disruption of polymerase-mediated amplification requires that blocker oligonucleotides not be elongated by the polymerase. Preferably, this is achieved through the use of blockers that are 3'-deoxyoligonucleotides, or oligonucleotides derivitised at the 3' position with other than a "free" hydroxyl group. For example, 3'-O-acetyl oligonucleotides are representative of a preferred class of blocker molecule.

Additionally, polymerase-mediated decomposition of the blocker oligonucleotides should be precluded. Preferably, such preclusion comprises either use of a polymerase lacking 5'-3' exonuclease activity, or use of modified blocker oligonucleotides having, for example, thioate bridges at the 5'-termini thereof that render the blocker molecule nuclease-resistant. Particular applications may not require such 5' modifications of the blocker. For example, if the blocker- and primer-binding sites overlap, thereby precluding binding of the primer (*e.g.*, with excess blocker), degradation of the blocker oligonucleotide will be substantially precluded. This is because the polymerase will not extend the primer toward, and through (in the 5'-3' direction) the blocker - a process that normally results in degradation of the hybridised blocker oligonucleotide.

A particularly preferred blocker/PCR embodiment, for purposes of the present invention and as implemented herein, comprises the use of peptide nucleic acid (PNA) oligomers as blocking oligonucleotides. Such PNA blocker oligomers are ideally suited, because they are neither decomposed nor extended by the polymerase.

In one embodiment of the method, the binding site of the blocking oligonucleotide is identical to, or overlaps with that of the primer and thereby hinders the hybridisation of the primer to its binding site. In a further preferred embodiment of the method, two or more such blocking oligonucleotides are used. In a particularly preferred embodiment, the hybridisation of one of the blocking oligonucleotides hinders the hybridisation of a forward primer, and the hybridisation of another of the probe (blocker) oligonucleotides hinders the hybridisation of a reverse primer that binds to the amplificate product of said forward primer.

In an alternative embodiment of the method, the blocking oligonucleotide hybridises to a location between the reverse and forward primer positions of the treated background DNA, thereby hindering the elongation of the primer oligonucleotides.

It is particularly preferred that the blocking oligonucleotides are present in at least 5 times the concentration of the primers.

In the fourth step of the method, the amplificates obtained during the third step of the method are analysed in order to ascertain the methylation status of the CpG dinucleotides prior to the treatment.

In embodiments where the amplificates were obtained by means of MSP amplification and/or blocking oligonucleotides, the presence or absence of an amplificate is in itself indicative of the methylation state of the CpG positions covered by the primers and or blocking oligonucleotide, according to the base sequences thereof.. All possible known molecular biological methods may be used for this detection, including, but not limited to gel electrophoresis, sequencing, liquid chromatography, hybridisations, real time PCR analysis or combinations thereof. This step of the method further acts as a qualitative control of the preceding steps.

In the fourth step of the method amplificates obtained by means of both standard and methylation specific PCR are further analysed in order to determine the CpG methylation status of the genomic DNA isolated in the first step of the method. This may be carried out by means of hybridization-based methods such as, but not limited to, array technology and probe based technologies as well as by means of techniques such as sequencing and template directed extension.

In one embodiment of the method, the amplicates synthesised in step three are subsequently hybridised to an array or a set of oligonucleotides and/or PNA probes. In this context, the hybridisation takes place in the following manner: the set of probes used during the hybridisation is preferably composed of at least 2 oligonucleotides or PNA-oligomers; in the process, the amplicates serve as probes which hybridise to oligonucleotides previously bonded to a solid phase; the non-hybridised fragments are subsequently removed; said oligonucleotides contain at least one base sequence having a length of at least 9 nucleotides which is reverse complementary or identical to a segment of the base sequences specified in the SEQ ID NO: 2 to SEQ ID NO: 5; and the segment comprises at least one CpG , TpG or CpA dinucleotide.

In a preferred embodiment, said dinucleotide is present in the central third of the oligomer. For example, wherein the oligomer comprises one CpG dinucleotide, said dinucleotide is preferably the fifth to ninth nucleotide from the 5'-end of a 13-mer. One oligonucleotide exists for the analysis of each CpG dinucleotide within the sequence according to SEQ ID NO: 1, and the equivalent positions within SEQ ID NOS: 2 to 5. Said oligonucleotides may also be present in the form of peptide nucleic acids. The non-hybridised amplicates are then removed. The hybridised amplicates are detected. In this context, it is preferred that labels attached to the amplicates are identifiable at each position of the solid phase at which an oligonucleotide sequence is located.

In yet a further embodiment of the method, the genomic methylation status of the CpG positions may be ascertained by means of oligonucleotide probes that are hybridised to the bisulfite treated DNA concurrently with the PCR amplification primers (wherein said primers may either be methylation specific or standard).

A particularly preferred embodiment of this method is the use of fluorescence-based Real Time Quantitative PCR (Heid et al., *Genome Res.* 6:986-994, 1996; *also see* United States Patent No. 6,331,393). There are two preferred embodiments of utilising this method. One embodiment, known as the TaqMan™ assay employs a dual-labelled fluorescent oligonucleotide probe. The TaqMan™ PCR reaction employs the use of a non-extendible interrogating oligonucleotide, called a TaqMan™ probe, which is designed to hybridise to a CpG-rich sequence located between the forward and reverse amplification primers. The

TaqMan™ probe further comprises a fluorescent “reporter moiety” and a “quencher moiety” covalently bound to linker moieties (e.g., phosphoramidites) attached to the nucleotides of the TaqMan™ oligonucleotide. Hybridised probes are displaced and broken down by the polymerase of the amplification reaction thereby leading to an increase in fluorescence. For analysis of methylation within nucleic acids subsequent to bisulfite treatment, it is required that the probe be methylation specific, as described in United States Patent No. 6,331,393, (hereby incorporated by reference in its entirety) also known as the MethyLight assay. The second preferred embodiment of this technology is the use of dual-probe technology (Lightcycler®), each carrying donor or recipient fluorescent moieties, hybridisation of two probes in proximity to each other is indicated by an increase or fluorescent amplification primers. Both these techniques may be adapted in a manner suitable for use with bisulfite treated DNA, and moreover for methylation analysis within CpG dinucleotides.

Also any combination of these probes or combinations of these probes with other known probes may be used.

In a further preferred embodiment of the method, the fourth step of the method comprises the use of template-directed oligonucleotide extension, such as MS-SNuPE as described by Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997. In said embodiment it is preferred that the methylation specific single nucleotide extension primer (MS-SNuPE preimer) is identical or complementary to a sequence at least nine but preferably no more than twenty five nucleotides in length of one or more of the sequences taken from the group of SEQ ID NO: 2 to SEQ ID NO: 5. However it is preferred to use fluorescently labelled nucleotides, instead of radiolabelled nucleotides.

In yet a further embodiment of the method, the fourth step of the method comprises sequencing and subsequent sequence analysis of the amplificate generated in the third step of the method (Sanger F., et al., *Proc Natl Acad Sci USA* 74:5463-5467, 1977).

Additional embodiments of the invention provide a method for the analysis of the methylation status of genomic DNA according to the invention (SEQ ID NO: 1) without the need for pretreatment.

In the *first step* of such additional embodiments, the genomic DNA sample is isolated from tissue or cellular sources. Preferably, such sources include cell lines, histological slides, body

fluids, or tissue embedded in paraffin. Extraction may be by means that are standard to one skilled in the art, including but not limited to the use of detergent lysates, sonification and vortexing with glass beads. Once the nucleic acids have been extracted, the genomic double-stranded DNA is used in the analysis.

In a preferred embodiment, the DNA may be cleaved prior to the treatment, and this may be by any means standard in the state of the art, in particular with methylation-sensitive restriction endonucleases.

In the *second step*, the DNA is then digested with one or more methylation sensitive restriction enzymes. The digestion is carried out such that hydrolysis of the DNA at the restriction site is informative of the methylation status of a specific CpG dinucleotide.

In the *third step*, which is optional but a preferred embodiment, the restriction fragments are amplified. This is preferably carried out using a polymerase chain reaction, and said amplicates may carry suitable detectable labels as discussed above, namely fluorophore labels, radionuclides and mass labels.

In the *final step* the amplicates are detected. The detection may be by any means standard in the art, for example, but not limited to, gel electrophoresis analysis, hybridisation analysis, incorporation of detectable tags within the PCR products, DNA array analysis, MALDI or ESI analysis.

The present invention enables prognosis of events which are disadvantageous to patients or individuals in which important genetic and/or epigenetic parameters within the PITX2 gene and its promoter or regulatory elements may be used as prognostic markers for breast cancer relapse or as 'adjuvant marker' for prediction of need of additional treatment besides of endocrine monotherapy. Said parameters obtained by means of the present invention may be compared to another set of genetic and/or epigenetic parameters, the differences serving as the basis for a prognosis of events which are disadvantageous to patients or individuals.

Specifically, the present invention provides for prognostic cancer relapse assays based on measurement of differential methylation of PITX2 CpG dinucleotide sequences. Preferred

gene sequences useful to measure such differential methylation are represented herein by SEQ ID NOS: 1 to 5. Typically, such assays involve obtaining a tissue sample from a test tissue, performing an assay to measure the methylation status of at least one of the inventive PITX2-specific CpG dinucleotide sequences derived from the tissue sample, relative to a control sample, and making a diagnosis or prognosis or prediction based thereon.

In particular preferred embodiments, inventive oligomers are used to assess PITX2 specific CpG dinucleotide methylation status, such as those based on SEQ ID NOS: 1 to 5, or arrays thereof, as well as a kit based thereon are useful for the prognosis of breast cancer relapse and/or the survival of a patient diagnosed with breast cancer, preferably under endocrine treatment since surgical removal of the tumor.

Moreover, an additional aspect of the present invention is a kit comprising, for example: a bisulfite-containing reagent as well as at least one oligonucleotide whose sequences in each case correspond, are complementary, or hybridise under stringent or highly stringent conditions to a 18-base long segment of the sequences SEQ ID NOS: 1 to 5. Said kit may further comprise instructions for carrying out and evaluating the described method. In a further preferred embodiment, said kit may further comprise standard reagents for performing a CpG position-specific methylation analysis, wherein said analysis comprises one or more of the following techniques: MS-SNuPE, MSP, MethyLight™, HeavyMethyl™, COBRA, and nucleic acid sequencing. However, a kit along the lines of the present invention can also contain only part of the aforementioned components.

Typical reagents (*e.g.*, as might be found in a typical COBRA-based kit) for COBRA analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); restriction enzyme and appropriate buffer; gene-hybridisation oligo; control hybridisation oligo; kinase labelling kit for oligo probe; and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kits (*e.g.*, precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

Typical reagents (*e.g.*, as might be found in a typical MethyLight®-based kit) for MethyLight® analysis may include, but are not limited to: PCR primers for specific gene (or

methylation-altered DNA sequence or CpG island); TaqMan® probes; optimised PCR buffers and deoxynucleotides; and Taq polymerase.

Typical reagents (*e.g.*, as might be found in a typical Ms-SNuPE-based kit) for Ms-SNuPE analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); optimised PCR buffers and deoxynucleotides; gel extraction kit; positive control primers; Ms-SNuPE primers for specific gene; reaction buffer (for the Ms-SNuPE reaction); and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kit (*e.g.*, precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

Typical reagents (*e.g.*, as might be found in a typical MSP-based kit) for MSP analysis may include, but are not limited to: methylated and unmethylated PCR primers for specific gene (or methylation-altered DNA sequence or CpG island), optimised PCR buffers and deoxynucleotides, and specific probes.

Definitions:

In the context of the present invention, the term “CpG island” refers to a contiguous region of genomic DNA that satisfies the criteria of (1) having a frequency of CpG dinucleotides corresponding to an “Observed/Expected Ratio” >0.6, and (2) having a “GC Content” >0.5. CpG islands are typically, but not always, between about 0.2 to about 1 kb in length.

In the context of the present invention, the term “methylation” refers to the presence or absence of 5-methylcytosine (“5-mCyt”) at one or a plurality of CpG dinucleotides within a DNA sequence.

In the context of the present invention the term “methylation state” is taken to mean the degree of methylation present in a nucleic acid of interest, this may be expressed in absolute or relative terms *i.e.* as a percentage or other numerical value or by comparison to another tissue and therein described as hypermethylated, hypomethylated or as having significantly similar or identical methylation status.

In the context of the present invention, the term "hemi-methylation" or "hemimethylation" refers to the methylation state of a palindromic CpG methylation site, where only a single cytosine in one of the two CpG dinucleotide sequences of the double stranded CpG methylation site is methylated (*e.g.*, 5'-NNC^MGNN-3' (top strand): 3'-NNGCNN-5' (bottom strand)).

In the context of the present invention, the term "hypermethylation" refers to the average methylation state corresponding to an *increased* presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

In the context of the present invention, the term "hypomethylation" refers to the average methylation state corresponding to a *decreased* presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

In the context of the present invention, the term "microarray" refers broadly to both "DNA microarrays," and 'DNA chip(s),' as recognised in the art, encompasses all art-recognised solid supports, and encompasses all methods for affixing nucleic acid molecules thereto or synthesis of nucleic acids thereon.

"Genetic parameters" are mutations and polymorphisms of genes and sequences further required for their regulation. To be designated as mutations are, in particular, insertions, deletions, point mutations, inversions and polymorphisms and, particularly preferred, SNPs (single nucleotide polymorphisms).

"Epigenetic modifications" or "epigenetic parameters" are modifications of DNA bases of genomic DNA and sequences further required for their regulation, in particular, cytosine methylations thereof. Further epigenetic parameters include, for example, the acetylation of histones which, however, cannot be directly analysed using the described method but which, in turn, correlate with the DNA methylation.

In the context of the present invention, the term "bisulfite reagent" refers to a reagent comprising bisulfite, disulfite, hydrogen sulfite or combinations thereof, useful as disclosed herein to distinguish between methylated and unmethylated CpG dinucleotide sequences.

In the context of the present invention, the term "Methylation assay" refers to any assay for determining the methylation state of one or more CpG dinucleotide sequences within a sequence of DNA.

In the context of the present invention, the term "MS.AP-PCR" (Methylation-Sensitive Arbitrarily-Primed Polymerase Chain Reaction) refers to the art-recognised technology that allows for a global scan of the genome using CG-rich primers to focus on the regions most likely to contain CpG dinucleotides, and described by Gonzalgo et al., *Cancer Research* 57:594-599, 1997.

In the context of the present invention, the term "MethyLight" refers to the art-recognised fluorescence-based real-time PCR technique described by Eads et al., *Cancer Res.* 59:2302-2306, 1999.

In the context of the present invention, the term "HeavyMethyl™" assay, in the embodiment thereof implemented herein, refers to a HeavyMethyl™ MethyLight assay, which is a variation of the MethyLight assay, wherein the MethyLight assay is combined with methylation specific *blocking* probes covering CpG positions between the amplification primers.

The term "Ms-SNuPE" (Methylation-sensitive Single Nucleotide Primer Extension) refers to the art-recognized assay described by Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997.

In the context of the present invention the term "MSP" (Methylation-specific PCR) refers to the art-recognised methylation assay described by Herman et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996, and by US Patent No. 5,786,146.

In the context of the present invention the term "COBRA" (Combined Bisulfite Restriction Analysis) refers to the art-recognized methylation assay described by Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997.

In the context of the present invention the term "hybridisation" is to be understood as a bond of an oligonucleotide to a complementary sequence along the lines of the Watson-Crick base pairings in the sample DNA, forming a duplex structure.

"Stringent hybridisation conditions," as defined herein, involve hybridising at 68°C in 5x SSC/5x Denhardt's solution/1.0% SDS, and washing in 0.2x SSC/0.1% SDS at room temperature, or involve the art-recognised equivalent thereof (e.g., conditions in which a hybridisation is carried out at 60°C in 2.5 x SSC buffer, followed by several washing steps at 37°C in a low buffer concentration, and remains stable). Moderately stringent conditions, as defined herein, involve including washing in 3x SSC at 42°C, or the art-recognised equivalent thereof. The parameters of salt concentration and temperature can be varied to achieve the optimal level of identity between the probe and the target nucleic acid. Guidance regarding

such conditions is available in the art, for example, by Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, N.Y.; and Ausubel et al. (eds.), 1995, Current Protocols in Molecular Biology, (John Wiley & Sons, N.Y.) at Unit 2.10.

"Background DNA" as used herein refers to any nucleic acids which originate from sources other than colon cells.

In the context of this application "survival" is meant to describe the time from diagnosis or start of treatment to an endpoint, which may be either the time of death (considering any reason for death or only death from breast cancer), or the time of recurrence of breast cancer (for example in form of metastases), which may be local or distant, or the time of occurrence of any breast cancer associated disease. Therefore "predicting the survival" is meant to comprise predicting the disease free survival, as well as the overall survival or any other consideration of time between diagnosis and endpoint of treatment.

Throughout this invention it is preferred that said survival is characterized as the disease free or the overall survival. It is especially preferred that survival is understood as disease free survival. Disease free survival is understood as absence of recurrence of cancer (local or distant) .

The terms "endocrine therapy" or "endocrine treatment" is meant to comprise any therapy, treatment or treatments targeting the estrogen receptor pathway or estrogen synthesis pathway or estrogen conversion pathway, which is involved in estrogen metabolism, production or secretion. Said treatments include, but are not limited to estrogen receptor modulators, estrogen receptor down-regulators, aromatase inhibitors, ovarian ablation, LHRH analogues and other centrally acting drugs influencing estrogen production.

The term "monotherapy" is used to explain that no other treatment is given in addition or to support said monotherapy.

In the context of the present invention the term "regulatory region" of a gene is taken to mean nucleotide sequences which affect the expression of a gene. Said regulatory regions may be located within, proximal or distal to said gene. Said regulatory regions include but are not limited to constitutive promoters, tissue-specific promoters, developmental-specific

promoters, inducible promoters and the like. Promoter regulatory elements may also include certain enhancer sequence elements that control transcriptional or translational efficiency of the gene.

In the context of the present invention the term "chemotherapy" is taken to mean the use of drugs or chemical substances to treat cancer. This definition excludes radiation therapy (treatment with high energy rays or particles), hormone therapy (treatment with hormones or hormone analogues (synthetic substitutes) and surgical treatment.

In the context of the present invention the term "adjuvant treatment" is taken to mean a therapy of a cancer patient immediately following an initial non chemotherapeutical therapy, e.g. surgery. In general, the purpose of an adjuvant therapy is to provide a significantly smaller risk of recurrences compared without the adjuvant therapy.

While the present invention has been described with specificity in accordance with certain of its preferred embodiments, the following examples and figures serve only to illustrate the invention and is not intended to limit the invention within the principles and scope of the broadest interpretations and equivalent configurations thereof.

In the sequence protocol and the Figures,

SEQ ID NO: 1 shows the sequence of the human gene PITX2,

SEQ ID NOS: 2 to 5 show chemically pretreated sequences of the gene PITX2,

SEQ ID NOS: 6 to 9 show the sequences of primers and probes according to PITX2 used in the example,

SEQ ID NOS: 10 to 12 show the sequences of primers and probes according to a control gene used in the example.

FIGURES

Figure 1 shows a preferred application of the method according to the invention. The X axis shows the tumour(s) mass, wherein the line '3' shows the limit of detectability. The Y-axis shows time. Accordingly said figure illustrates a simplified model of endocrine treatment of

an Stage 1-3 breast tumour wherein primary treatment was surgery (at point 1), followed by adjuvant therapy with Tamoxifen, as an example for an endocrine treatment. In a first scenario a patient without relapse during endocrine treatment (4) is shown as remaining below the limit of detectability for the duration of the observation. A patient with relapse of the cancer (5) has a period of disease free survival (2) followed by relapse when the carcinoma mass reaches the level of detectability.

Figure 2 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of the PITX2 gene by means of Real-Time methylation specific probe analysis. The lower plot shows the proportion of disease free patients in the population with above median methylation levels, the upper plot shows the proportion of disease free patients in the population with below median methylation levels. The X axis shows the disease free survival times of the patients in months, and the Y- axis shows the proportion of disease free survival patients.

EXAMPLE

Real time Quantitative methylation analysis

Genomic DNA was analyzed using the Real Time PCR technique after bisulfite conversion. In this analysis four oligonucleotides were used in each reaction. Two non methylation specific PCR primers were used to amplify a segment of the treated genomic DNA containing a methylation variable oligonucleotide probe binding site. Two oligonucleotide probes competitively hybridise to the binding site, one specific for the methylated version of the binding site, the other specific to the unmethylated version of the binding site. Accordingly, one of the probes comprises a CpG at the methylation variable position (i.e. anneals to methylated bisulphite treated sites) and the other comprises a TpG at said position (i.e. anneals to unmethylated bisulphite treated sites). Each species of probe is labelled with a 5' fluorescent reporter dye and a 3' quencher dye wherein the CpG and TpG oligonucleotides are labelled with different dyes.

The reactions are calibrated by reference to DNA standards of known methylation levels in order to quantify the levels of methylation within the sample. The DNA standards were composed of bisulfite treated phi29 amplified genomic DNA (i.e. unmethylated), and/or phi29 amplified genomic DNA treated with Sss1 methylase enzyme (thereby methylating each CpG position in the sample), which is then treated with bisulfite solution. Seven

different reference standards were used with 0%, (i.e. phi29 amplified genomic DNA only), 5%, 10%, 25%, 50%, 75% and 100% (i.e. phi29 Sss1 treated genomic only).

The amount of sample DNA amplified is quantified by reference to the gene (β -actin (ACTB)) to normalize for input DNA. For standardization the primers and the probe for analysis of the ACTB gene lack CpG dinucleotides so that amplification is possible regardless of methylation levels. As there are no methylation variable positions, only one probe oligonucleotide is required.

The following oligonucleotides were used in the reaction:

Primer: TGGTGATGGAGGAGGTTTAGTAAGT (SEQ ID NO: 10)

Primer: AACCAATAAAACCTACTCCTCCCTTAA (SEQ ID NO: 11)

Probe: 6FAM-ACCACCACCCAACACACAATAACAAACACA-TAMRA or Dabcyl (SEQ ID NO: 12)

The extent of methylation at a specific locus was determined by the following formula:

$$\text{methylation rate} = 100 * I(\text{CG}) / (I(\text{CG}) + I(\text{TG}))$$

(I = Intensity of the fluorescence of CG-probe or TG-probe)

Gene PITX2 (SEQ ID 1, und 2-5)

Primers:

PITX2R02: GTAGGGGAGGGAAGTAGATGTT (SEQ ID NO: 6)

PITX2Q02: TTCTAATCCTCCTTTCCACAATAA (SEQ ID NO: 7)

Amplificate length : 143 bp

Probes:

PITX2cg1: FAM-AGTCGGAGTCGGGAGAGCGA-Darquencher (SEQ ID NO: 8)

PITX2tg1: YAKIMA YELLOW-AGTTGGAGTTGGGAGAGTGAAAGGAGA-Darquencher (SEQ ID NO: 9)

PCR components: 3 mM MgCl₂ buffer, 10x buffer, Hotstart TAQ

Program (45 cycles): 95 °C, 10 min

95 °C, 15 sec

62 °C, 1 min

Figure 2 shows the Kaplan-Meier estimated disease-free survival curves for a CpG position of the PITX2 gene by means of Real-Time methylation specific probe analysis. The lower plot shows the proportion of disease free patients in the population with above median methylation levels, the upper plot shows the proportion of disease free patients in the population with below median methylation levels. The X axis shows the disease free survival times of the patients in months, and the Y- axis shows the proportion of disease free survival patients. The p-value (probability that the observed distribution occurred by chance) was calculated as 0.0031, thereby confirming the data obtained by means of array analysis.

CLAIMS:

1. A method for characterising a cell proliferative disorder of the breast tissues and/or **predicting the survival** of a patient diagnosed with said disorder, comprising the steps of:
 - (a) obtaining one or more biological samples from the patient; and
 - (b) detecting the level of expression of a polypeptide expressed from the PITX2 gene
2. The method according to claim 1 further comprising
 - (c) determining therefrom the survival of said patient, characteristics of said cell proliferative disorder, and/or prognosis of said patient
3. The method according to claim 1 further comprising
 - (d) determining a suitable treatment regimen for the subject
4. The method of claim 1, wherein said patient is characterized by being subject to adjuvant endocrine therapy comprising one or more treatments which target the estrogen receptor pathway or are involved in estrogen metabolism, production or secretion.
5. The method of claim 1, wherein said breast cell proliferative disorders are taken from the group comprising ductal carcinoma *in situ*, invasive ductal carcinoma, invasive lobular carcinoma, lobular carcinoma *in situ*, comedocarcinoma, inflammatory carcinoma, mucinous carcinoma, scirrhous carcinoma, colloid carcinoma, tubular carcinoma, medullary carcinoma, metaplastic carcinoma, and papillary carcinoma and papillary carcinoma *in situ*, undifferentiated or anaplastic carcinoma and Paget's disease of the breast.
6. The method of claim 1, wherein said detection is afforded by performing an immunoassay, in particular by an ELISA.
7. The method of claim 6, wherein said immunoassay is a radioimmunoassay.
8. A method of predicting survival of a patient diagnosed with a cell proliferative disorder of the breast, comprising the steps of:

- a) obtaining one or more biological test samples from said patient; and
 - b) contacting said sample with an antibody immunoreactive with the PITX2 polypeptide to form an immunocomplex;
 - c) detecting said immunocomplex;
 - d) comparing the quantity of said immunocomplex to the quantity of immunocomplex formed under identical conditions with the same antibody and a control sample from one or more patients with a known prognosis; wherein a decrease in quantity of said immunocomplex in the sample from said subject relative to said control sample is indicative of a bad prognosis.
9. The method of claim 8, wherein said immunocomplex is detected in a Western blot assay.
10. The method of claim 8, wherein said immunocomplex is detected in an ELISA.
11. The method of claim 1, wherein said detection is afforded by expression analysis.
12. The method of claim 11, comprising detecting the presence or absence of mRNA encoding a PITX2 polypeptide in a sample from a patient, wherein a decreased concentration of said mRNA below the concentration determined for an individual known to have a good prognosis indicates a bad prognosis.
13. The method of claim 11, comprising the steps of:
- a) providing a polynucleotide probe which specifically hybridises or is identical to a polynucleotide consisting of SEQ ID NO: 1,
 - (b) incubating said sample with said polynucleotide probe under high stringency conditions to form a specific hybridisation complex between an mRNA and said probe;
 - (c) detecting said hybridisation complex.
14. The method according to claim 13 wherein the detecting step further comprises the steps of:
- a) producing a cDNA from mRNA in the sample;
 - b) providing two oligonucleotides which specifically hybridise to regions flanking a segment of the cDNA;
 - c) performing a polymerase chain reaction on the cDNA of step a) using the

- oligonucleotides of step b) as primers to amplify the cDNA segment; and
d) detecting the amplified cDNA segment.
15. Use of a polypeptide expressed from the PITX2 gene for differentiating or distinguishing between patients diagnosed with breast cancer, which have a good survival prognosis and patients which have a bad survival prognosis.
 16. Use of a polypeptide expressed from the PITX2 gene for prediction of survival of a patient diagnosed with a cell proliferative disorder of the breast.
 17. The method of claim 1 wherein said detection comprises determining the genetic parameters of the gene PITX2, its promoter and/or regulatory elements.
 18. The method of claim 1 wherein said detection comprises determining the epigenetic parameters of the gene PITX2, its promoter and/or regulatory elements.
 19. The method of claim 1, wherein said detection comprises determining the **methylation** status of one or more CpG positions of a target nucleic acid within the gene PITX2, its promoter and/or regulatory elements, in particular through the methylation analysis of a genomic DNA sequence according to SEQ ID NO: 1.
 20. The method of claim 19, wherein the methylation analysis is afforded by contacting said target nucleic acid with one or more agents that convert cytosine bases that are unmethylated at the 5'-position thereof to a base that is detectably dissimilar to cytosine in terms of hybridization properties.
 21. The method of claim 20, wherein contacting said target nucleic acids with one or more agents comprises use of a solution selected from the group consisting of bisulfite, hydrogen sulfite, disulfite, and combinations thereof.
 22. A nucleic acid comprising a sequence at least 18 bases in length of a segment of the chemically pretreated genomic DNA according to one of the sequences taken from the group comprising SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto.

23. An oligomer, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridises under moderately stringent or stringent conditions to a pretreated genomic DNA according to one of the SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto.
24. The oligomer as recited in Claim 23; wherein the base sequence includes at least one CpG, tpG or Cpa dinucleotide.
25. The oligomer as recited in Claim 24; characterised in that the cytosine of the CpG dinucleotide is located approximately in the middle third of the oligomer.
26. A set of oligomers, comprising at least two oligomers according to any of claims 23 to 25.
27. A set of at least two oligonucleotides as recited in Claims 23 to 26, which can be used as primer oligonucleotides for the amplification of DNA sequences of one of SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto.
28. Use of a set of oligomer probes comprising at least ten of the oligomers according to any of claims 24 to 27 for detecting the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) within one of the sequences according to SEQ ID NO: 1, and sequences complementary thereto.
29. A method for manufacturing an arrangement of different oligomers (array) fixed to a carrier material for analysing diseases associated with the methylation state of the CpG dinucleotides of one of SEQ ID NO: 1, and sequences complementary thereto wherein at least one oligomer according to any of the claims 23 through 27 is coupled to a solid phase.
30. A composition of matter comprising the following:
 - a nucleic acid comprising a sequence at least 18 bases in length of a segment of the chemically pretreated genomic DNA according to one of the sequences taken from the group comprising SEQ ID NO: 1 to SEQ ID NO: 5 and sequences complementary

thereto, and

- a buffer comprising at least one of the following substances: 1 to 5 mM Magnesium Chloride, 100-500 μ M dNTP, 0.5-5 units of taq polymerase, an oligomer, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridises under moderately stringent or stringent conditions to a pretreated genomic DNA according to one of the SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto.

35. Use of the gene PITX2, its promoter and/or regulatory elements for detecting the survival of patients diagnosed with a cell proliferative disease.

36. A method for detecting the survival of patients diagnosed with a cell proliferative disease according to claim 19, comprising:

- a) obtaining, from a subject, a biological sample having subject genomic DNA;
- b) treating the genomic DNA, or a fragment thereof, with one or more reagents to convert 5-position unmethylated cytosine bases to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridisation properties;
- c) contacting the treated genomic DNA, or the treated fragment thereof, with an amplification enzyme and at least two primers comprising, in each case a contiguous sequence at least 18 nucleotides in length that is complementary to, or hybridises under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NOS: 2 to 5, and complements thereof, wherein the treated DNA or a fragment thereof is either amplified to produce one or more amplicates, or is not amplified; and
- d) determining, based on the presence or absence of, or on a property of said amplicate, the methylation state of at least one CpG dinucleotide sequence of SEQ ID NO: 1, or an average, or a value reflecting an average methylation state of a plurality of CpG dinucleotide sequences of SEQ ID NO: 1.

37. A method for detecting the survival of patients diagnosed with a cell proliferative disease according to claim 19, comprising the following steps of

- a) obtaining, from a subject, a biological sample having subject genomic DNA;
- b) treating the genomic DNA, or a fragment thereof, with one or more reagents to

convert 5-position unmethylated cytosine bases to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridisation properties;

c) amplifying one or more fragments of the treated DNA such that only DNA originating from colon or colon cell proliferative disorder cells are amplified

d) detecting the amplicates or characteristics thereof and thereby deducing on the presence or absence of a colon cell proliferative disorder.

38. The method of one of claims 36 or 37, wherein in step a) the biological sample obtained from the subject is selected from the group consisting of histological slides, biopsies, paraffin-embedded tissue, bodily fluids, serum, plasma, stool, urine, blood, nipple aspirate and combinations thereof.

39. A method for detecting a colon cell proliferative disorder according to claim 17, comprising:

a) obtaining, from a subject, a biological sample having subject genomic DNA;

b) extracting the genomic DNA;

c) contacting the genomic DNA, or a fragment thereof, comprising SEQ ID NO:1 or a sequence that hybridises under stringent conditions to SEQ ID NO:1, with one or more methylation-sensitive restriction enzymes, wherein the genomic DNA is either digested thereby to produce digestion fragments, or is not digested thereby; and

d) determining, based on a presence or absence of, or on property of at least one such fragment, the methylation state of at least one CpG dinucleotide sequence of SEQ ID NO: 1, or an average, or a value reflecting an average methylation state of a plurality of CpG dinucleotide sequences of SEQ ID NO: 1, whereby at least one of detecting the prostate cell proliferative disorder, or distinguishing between a transitional and a peripheral zone of origin of the prostate cell proliferative disorder is, at least in part, afforded.

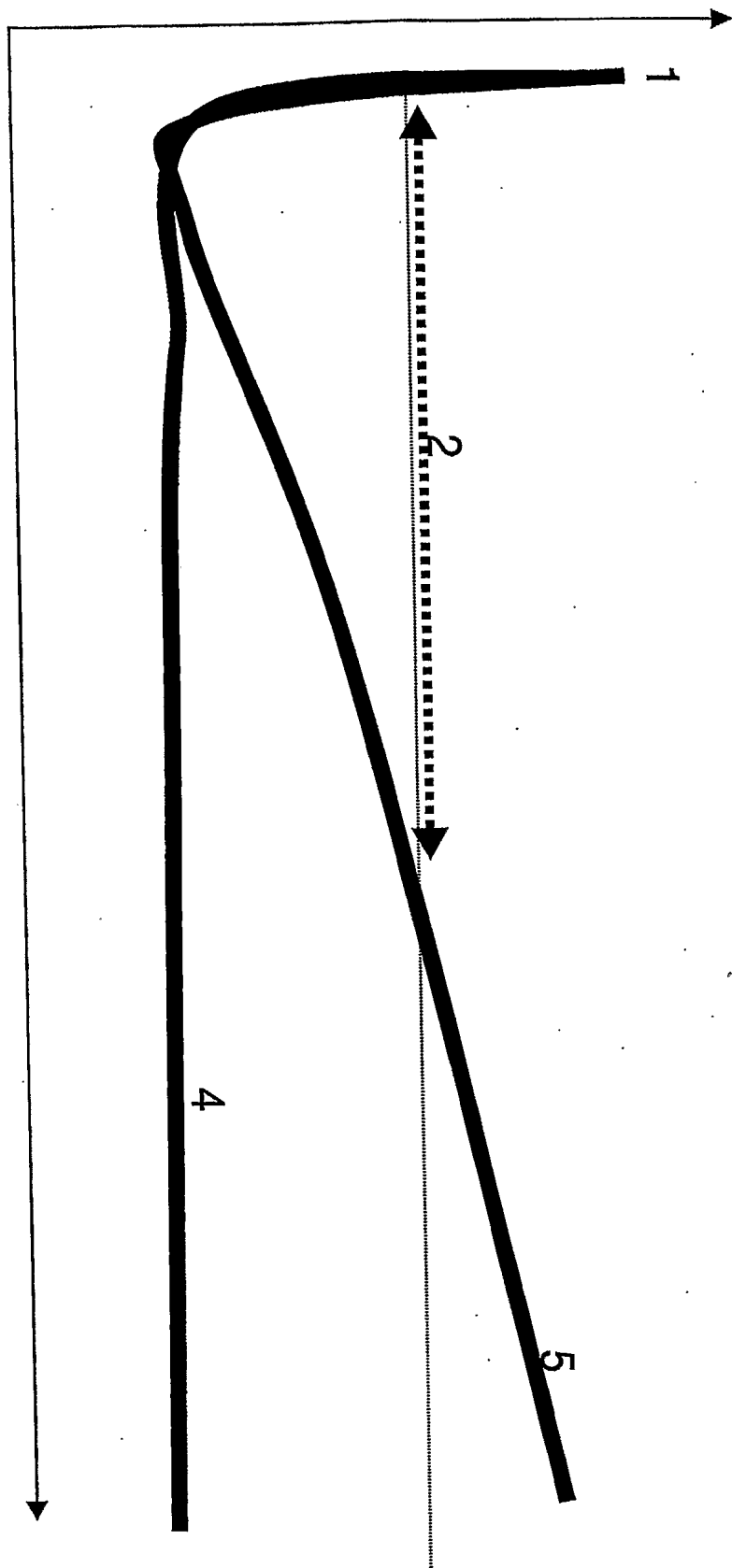
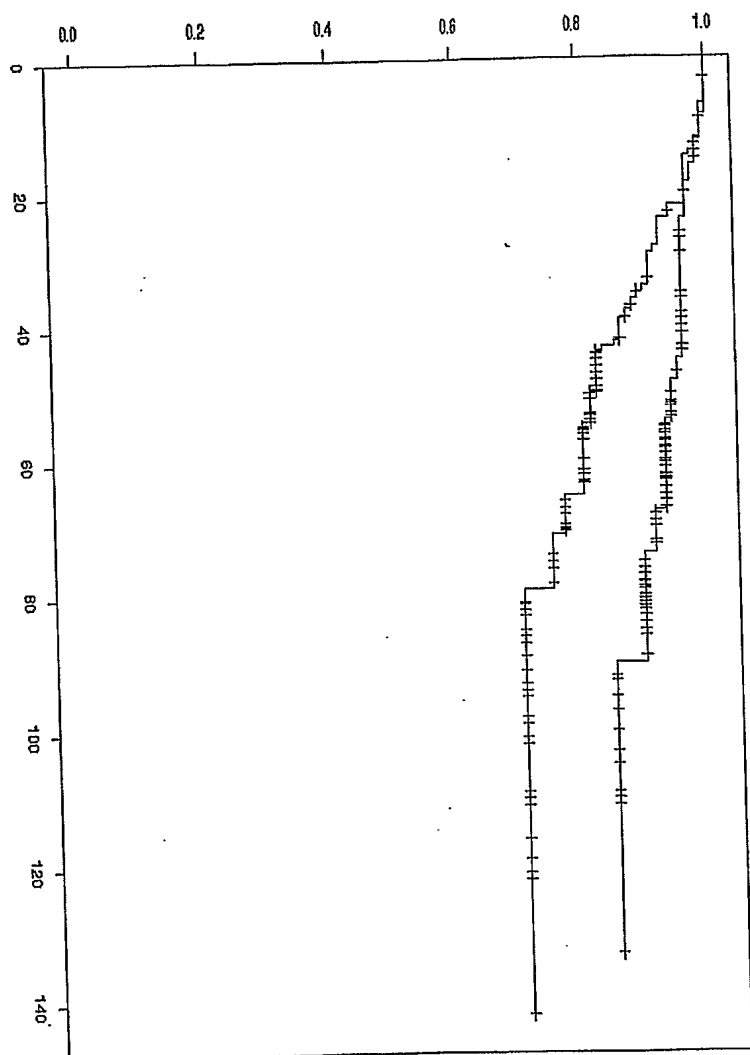


Figure 1

Figure 2



Sequence listing

<110> Epigenomics AG

<120> PITX2 - a marker to predict survival of patients diagnosed with breast cell proliferative disease

<160> 12

<210> 1

<211> 9001

<212> DNA

<213> Homo Sapiens

<400> 1

agctgatgga	cttgctaaat	ttctttcttc	ttttttcttt	ttcatattat	ttgctagcca	60
taatggaatc	ctctagggtt	aagccaaaga	aaaattggag	agacaaaatt	agatttttga	120
gcccttttcc	cccccgggaa	tgcccttttt	tttcttttta	gtttctgatg	aatggctatc	180
atltatttct	accaaattta	aataaggact	gctgccttgt	atgtttaact	atggcaggag	240
aggggaactg	tttgttttag	aagcagtgc	tgagatgtcc	tgccaaggtt	agtgcagag	300
gaggggagaa	agaatccaga	ccaatttgta	tgagatgtcc	tttactccca	tgaaataaaa	360
cacatttgtt	tcataattgc	tgaaaagtaa	aacaataata	ttgtacgaaa	tggtatacac	420
agggttaggt	gtacatagca	gtttcagaaa	catcattgca	tcaccagag	aaactattct	480
aaaactgata	ttcacacatt	ttttataata	ataataatat	gttagaaaca	tacagtgtgg	540
catttagtat	atacactccc	ttgctcgcga	gcgaaaaatc	ctaactcgct	ctgtataaca	600
tgctttatgt	ttaaagcctaa	cctttaaaaa	cactgtttgt	atattactaa	caactgcttt	660
tataaaatta	atttgacatt	tcgatatata	tacatccttt	cagtcattta	aatgttaaca	720
atgctaaact	taaaaaataa	caagcttata	gtaatgttaa	aatgtcatat	ccagtcacac	780
atgtgtttgt	gtatgtgtcc	ttgcaactgt	tagaaatact	tgtagtgaaa	gatgtcagac	840
actgaggaca	tccctttgaa	atcaaaggag	ctctctcttt	gattcagtgg	tttctttttc	900
tctatatagc	ttctctttct	ctccctttct	ttagtgccta	cgaccttcta	gcataattcc	960
cagtcctttca	agggcgggag	tgcccccatt	ggcaagggtc	taggatcccg	gcgctgtggg	1020
tgccggtcac	acgggcgggt	ccactgcata	ctggcaagca	ctcagggttg	agggcgggtt	1080
ctgcacgctg	gcgtagccga	agctggagtg	ctgcttttgt	ttcagttcca	ggctggccag	1140
gctcgagtta	cacgtgtccc	tataaacata	cggaggagtc	ggcggcgctg	aaggacaggg	1200
agggcgtcgg	accgcggaat	tcagcgacgg	gctactcagg	ttgttcaagt	tattcaggct	1260
gttgagactg	gagcccgggg	cgctgtcac	tgctgagggc	accatgctgg	acgacatgct	1320
catggacgag	atagagttgg	gtggggaaaa	catgctctgt	gatgacaggg	gggtgacgtt	1380
catagagttg	ttgtagggaat	agctcttggg	ggatagggag	gcggatgtaa	ggcccttggc	1440
gaactgcggc	ccgaagccat	tcttgcatag	ctcggcctgc	tggtttgcgt	ccctctttct	1500
ccatttggcc	cgacgattct	tgaacccaaac	ctggggggcg	ttggggcaag	ggagcaaaaca	1560
gatgccacag	tgagatttac	taaaactttc	atcggaggcc	aacccccgcc	ttcccccgac	1620
acacacgcta	gcgcactcac	acacccctgg	ctcgcttcac	tgacccggcc	tgaccccaaa	1680
gataccaggg	ccagcttttca	gttactggcc	cgggtctcca	ccaagcgag	gagacctggt	1740
ctgctctggc	ctgcgagctg	ggactcggag	ctacgccaca	aacctcagcc	gaacgcagtg	1800
agacctgcgg	acggttttgat	cactcagcca	ggcgtttctc	caggtccaaa	aacacttaat	1860
gtaaaaacaaa	cgcggggcag	caggcttttc	caacccttcc	cggggcacct	tgcaaaacttg	1920
cttccattcc	aaagccacag	acccacggat	gaggagaagg	ggctggaagg	gcactagagg	1980
atcgctcttt	ctcccacgca	attcctccct	tccttccctg	acctccactg	tcgtcccca	2040
ccccctggta	cgtgctccct	taacagggag	taggccgcca	acactctttc	tcgcctagca	2100
aaacaaccaa	ataaagagca	aaagaccacc	tcttcgtcag	ctcgtaact	ccaggagctt	2160
ggcatattaa	actccgggaa	cccggaaagg	gtagttttgg	agattccccc	ttctttcgct	2220
ctgctctctt	tttaccctaa	gcccaccaca	ggcctgtccg	cgcccgaggc	ccagccgggt	2280
cgttttggct	tgacggcggc	caccacgggc	ggccgggttc	caccctgtgc	cggttgccca	2340
gccgcaacc	cgatcccaat	ccacatcggg	cctccctgtc	gcccagacg	gggctttttg	2400
tgtattggag	agaggcctgg	cctgagatat	ccgagctgac	accagtgtat	tttcacatta	2460
cacatctccg	ccggggccag	ccgtgtaatc	cgctttttct	ctttttcctt	tcattcttga	2520
tttctttttt	atcccccttc	ctctttgcac	ccgactgcta	taaaaagcac	gcctcactcc	2580
cacttggttc	gacaagcagc	cgccctggaa	ggagaggcag	ctgcaaggag	agcccagcgc	2640
cgcggtctaca	aagcactagg	gtggagctgc	ggaatagcgg	gcgggggtgg	agggcgcttt	2700
cgaaggatcc	cagaaaaccc	atagactctg	tctttaatta	cttgccattt	ctaccctagg	2760
ccatctaaac	tttgctcagg	cgagaagagt	acgtgagagg	cccgttccct	tgatgtgcaa	2820
gagagcta	gaaagactga	ccttgctcaa	aaccacgccc	cccaggaccc	agctctggct	2880
ctggacagtt	aaactaaaac	catttttcaa	ttctttcccg	ccttttatcc	accagcatag	2940
cctcatgctt	tgacacaaat	ccaccagag	agtgtcttca	ttccctctga	tttgggagag	3000
cattttgggt	tttattcttt	ttatcgttgt	tttcttcttt	ttgtttgctc	tgctctaaac	3060
gggggcttta	ttttttctac	ccagagcact	taattttttt	tttttaacag	caaagcctct	3120
ggatgccgct	tgatttgctt	gattctgttt	tctgcttcca	gaatcctaac	aaatttggaa	3180
tcttccaccg	accagcataa	accaggacgt	tgctattggg	ttatttattt	gagctcattt	3240
						3300

ttgccaatoc	ataaagtaca	gattttgctac	aaagttaagg	taagcccttt	ttacaaaact	3360
atgattataa	tttagaagag	ggggtgtgag	tttcaatttc	cagagttcaa	ctcctgagag	3420
aagataaata	aaccaagcag	aaaagtcttt	ctttcttttt	tctttctcct	tctaagagga	3480
ctagtagttg	tgtattaaaa	ctttgtctcc	ggagatcaca	aaactagga	atagggtgtg	3540
tgggagagac	ctgaatggcc	gaaacaaccg	taaagaaggt	gtaagaagcg	cgagcccagg	3600
agggaaaaag	ctggggccagg	gccgggacaa	aggtttccca	gggagggcca	actcttccgt	3660
gtctctggcg	ggttttcctt	gttaaaggct	cacagggttg	agcctgttcg	cggctcttgg	3720
cctggtaggg	atttttattag	ctctgtctctg	gcaactgcaa	gccaggaaca	caatgtcctg	3780
tgcaggggat	tgcccatgca	gccagctcgt	tgagatcgcg	ggatggcggg	gcagtggagc	3840
gggtgcgctc	tgggagcctg	agccagggcg	gcagtctctg	cggcctcgga	gagggaaactg	3900
taatctcgca	accaggccgc	cgcgaggcct	tctgcctttg	caaagctgcg	ccccaccggc	3960
gccttcccag	gcggcgctgc	cttccacatt	ctctcctggg	ctacttggcc	tgtacctcca	4020
caacatcctc	ccccatccc	tcccagactc	cgtgtctggc	cctaccggga	ctcgggcttc	4080
cgtaagggtg	gtccacacag	cgatttcttc	cggtgtggac	atgtccgggt	agcgggttct	4140
ctggaaagtg	gcctccagct	cctggagctg	ctggctggta	aagtgagtc	gctgccgcct	4200
ttgcgccttc	ttcttagacg	ggtcctcggc	gcccacgtcc	tcattcttcc	cctgctgggt	4260
tttatctttt	tctgaaaacg	aaacacacac	actttcccg	cagcatgccc	acctgcaacg	4320
oggagcccaa	ctggaccggc	ggcagaagcc	gtggaagagc	tgggctgcct	ggcgctggag	4380
gaggggtgcg	gcggcggtc	ggggccgcga	ggagcgctgc	gcctgtgggg	tgtgcaggcg	4440
caagtgtggg	tgtccgcgcc	ccatttccct	ccctccccc	gcgcgcgacg	ttttattttac	4500
atgttttatct	cactgcagcg	gcacattcac	ttttatagcc	tgtgctttca	agtatatatta	4560
tacacctctg	cgcagacaca	ccaaatctcc	tgggacgcgc	acacgcgcgt	ggttttacaga	4620
ccccctcccc	cctcgcagaa	agctcagatt	tccatgcggt	ttgggaaggc	taggaaaaga	4680
tgtggggatt	cggttgggca	ccgaagtctg	ccggcccttt	ccccaaaaaa	aaaaaaaaaat	4740
gcctcttcgc	gaagggcatt	tctgagtggt	ttcaggcaat	ttcctaacga	gtggagctcc	4800
tggggagctg	aaagccgaga	ggaaaacagg	gacagaggtc	ggcgccctct	gaaggctctc	4860
gaatcaagat	gctgggattt	ttgtgaccca	ggaaaacagaa	gggaggccag	ggtagcaata	4920
gagagggcgg	cagaattgct	cgcgccctta	gcgccccagg	agccggggcg	gtcgagggag	4980
aactaaaggg	atgcggggta	gtcaaaaattc	cggctcccgg	aagttctgcg	gggagccagg	5040
cgaacgacca	ctcccaccac	gcctcccccc	ggaggggctg	acttccctgg	ggcgagaggg	5100
agcgggtggc	gcagagcagc	tgagcgggaa	tgtctgcagg	gcggcgcgcc	gccttacctg	5160
cggcctccgg	gctggagggtg	tcggagatgg	tgtgcacctc	cagcctgtgc	ttggaggagt	5220
ccagcgaccg	gggctgaccg	ggagccagaa	ccgaagccat	ggctaaccgg	tggggatggt	5280
gacaggaaga	tgaggagacg	gccgacagct	tgggtccccg	tgctcgggtg	tccaagttaa	5340
gcgggccttt	catgcagttc	atggacgagg	gagcgcgacg	ctctactagt	ctctggctac	5400
tgcccccgcc	agcccccgta	gccgcgctg	cccgtccggg	gtcgcgctct	agggcgggag	5460
tttccccgct	gcgggggagag	ccagggggacg	caacccccgc	cgagttctca	agccaagctg	5520
ccccgctctc	ctccggaagg	ctcaagcgaa	aaagtccgga	gacggaaagt	cagcgggcaa	5580
acgaagacat	gggatgtggg	cagaagggca	ccactcagag	cgtcttttag	gagcaggctt	5640
ccaagctcca	agagtgggca	agagccccct	aagaccctct	tcttctctcc	ctccctcccc	5700
caagaacccc	tccaataagg	aaagctaacg	ccgaccgcgc	tctgcccggc	ccccccccac	5760
gcggcagccc	tgacagagaa	gtgtcaagag	tgacagggac	aggtaggtga	tattagatcc	5820
cctgcggcgg	cagcagccgc	tgcagccacg	acgcggccct	ctgagcgcac	cctccgcaac	5880
gcgcacacgc	acaccctctg	ggcggtcgaa	caggagcccg	gccttgccgc	agctcagctc	5940
caggcacccc	ggcgagcgac	ggaccagatc	tgcggctccg	cgcttccctg	ttggcctaac	6000
atcttaaaac	cagagggcggg	cttccctggtg	ccgagacgtc	actccgcgcg	ggccctcccc	6060
agccctctcc	gcctccgcct	cctcccagac	ccttctccgg	gtgcgactga	cgtggctccg	6120
caccaatcag	gacgccccga	gccgcggtgg	agggactgtc	ctgcctgcac	ctatcagcag	6180
tgcggggcgg	ggctactgcc	tcgcccgtcg	cactgggtct	acacaggcaa	gctcccggga	6240
attcagctcc	tgcccagccc	aaggcgatcc	ggcttttagt	acgaacccaa	aggtgaagag	6300
atgaggctag	gagtcgaagg	cttggggagaa	gagagtggaa	tgggtcaagaa	gagaaaggta	6360
caaggatcaa	caagacaccc	actctttgtg	tctcactaca	tccattttcca	atccccacc	6420
ccatataaaa	aggagacacg	ttacttaaaaa	ctagaaaaatt	tgaaaaaacag	caacaaatca	6480
cctctccgat	cttaaatttt	ccaaacagcc	tgtcaagtga	atgctgcgct	aatctgaaga	6540
agctttaatt	gcaaagaaga	cagagccctg	aaaaggcagg	ctaataaatt	agaaatcgag	6600
aagcaaattg	acccgtcaaa	agaaaattac	cttgacttta	aacgaacaac	tgtttggtgg	6660
ttcactcttg	atttatacaa	gaataaaaaa	tcgcctcaga	tcacgttctc	tgtgatgctt	6720
attagctccc	agacagaaaa	cacacaatag	aagagaaacc	ctaaccagc	gttttcaaaa	6780
tgtgaaagc	ttatccattc	tacttaacgt	tgattaagac	acatatccca	gatctttcaa	6840
attccttgta	cactgtatta	agctcgtcct	aacccgagag	agccacgctt	taaattcgac	6900
tctcttgttt	actttattat	caatcagatt	taaattccata	aagcctgtag	aatcaacaac	6960
cttgagctaa	ttatatatga	aatatgcctt	aatgaatttc	catacaatta	agaatgttgc	7020
caaataacca	atltcaaggga	taatttttaa	cagtcatttt	cttttcccag	tgagctcaag	7080
gctgtcttga	gccattaaag	tccaagcagg	cagaaggggt	gtgtgtgagc	taagggcgaa	7140
aagcctagaa	ctgcgctcaa	ctagcaaaaag	caaaaacctta	tttatataaaa	acaaaaaaa	7200
tcacctttgg	agacatcaac	tcttttatagc	actgttttcca	agcaaattta	atltccaaag	7260
aaatttaaaga	aagaaatcca	aacatattca	aaataatttt	tgaaagtcct	tttgtcccc	7320
agcataggtc	agctgggagag	gacaaactaa	tctcctctgg	gtttctgcat	gggcgattgt	7380
tttactatgg	agttagtgtt	atcatctctg	aatgtgtatt	tgtttgacat	tacagtcaat	7440
gatttgcaat	gccagcatga	agtatcttta	aaacactccc	tccttgteet	tgttcaacag	7500
attgggaaac	ttattogatg	tggaaacaaag	tggatgaagc	agactacaaa	tatatattgca	7560

acttatgtgt	ctctcctttg	ccctgaccac	ccccaaaccc	tatctgcaac	tcctccccat	7620
tttaaacttg	cagtcctaaag	acgcacatga	gaattgtttt	tcagtcctttc	ttcaccagta	7680
tcacccact	ttaagaataa	tttagctgca	agggaggaat	ttcttcataag	taagctttta	7740
atcagcattt	ctgcttttaa	ccttttatcc	cactttaccc	cattccacac	atacagacac	7800
ctgctcagag	taaaacacat	cctcatgtga	caggtctgca	ttagctgagg	ctcatacatc	7860
cagctatatt	aggctctgca	atcttatcac	taaattatac	acattacact	agcagcctgt	7920
tggttaaagaa	ggttaaatta	atctacattc	tgtctattat	ctggtgctta	aatgacgcat	7980
tttatcccg	agatttggcg	gagaatctcc	ttctcagacc	ccacagcgtt	tcactgaaga	8040
caatgctctt	acatttgtag	tggtttttta	tctgataaga	ctctaatttg	cttaagtctt	8100
ttaaataagg	gttttaaatg	tttctagccg	ttttcttatt	gaatttcctc	taattcccc	8160
aagatcataa	agtatatgtg	ttaaagtaaat	atctctccc	attgcaactgc	cagccgatga	8220
cctataacta	agtcaataag	aatccagctc	ttttctgctg	aatgtgttta	ctaatacatat	8280
tccagttttct	tcttttaaac	ctcagaatag	ctgtgggtccc	cacaatacca	tgccccttaa	8340
agcttcattt	catgaaggga	ctccatcaca	ttaaagaatg	aaaaaaatct	ccactgtagt	8400
tagtatacat	agccctcac	tccttgtttt	tcaagattca	aaccccagag	ctgcaaatat	8460
ttttggaagc	ttgggtgtta	atgccttatt	ttagaaagcc	gagaagcccc	acagagccat	8520
atagattttct	aaacccatct	ctcataaacc	cacagaattt	tgataaaagc	tctggtggct	8580
ccactctacc	gatggaactt	tcatacagac	aaatatacat	gtatgaagga	cctcaatcag	8640
cctccaaagt	ggttgaaaaa	cccaagggca	cgtgactgat	cctcatagt	ccaacgtgtg	8700
cgagatgttg	gaagcactgg	ggatcagcag	cagcctagat	gcctaaaaag	ataaggtgtc	8760
ctaattttgtg	tggaacccatt	gaagtcaagt	gggtgaataaa	gacaattatc	tagataattc	8820
agattaaagt	aaaagcaaaa	ccatatctat	ttgtatatat	atattcacat	ccattttata	8880
ttacagacac	acacacgcac	acacacactg	gctctgtaaa	caactgactc	aaagttagga	8940
ttttctttgc	atctttccag	caggagtctc	aacattctcc	taatctccta	atcactttac	9000
a						9001

<210> 2

<211> 9001

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 2

agttgatgga	tttggttaaat	tttttttttt	tttttttttt	tttatattat	ttgttagtta	60
taatggaatt	tttttaggttt	aagttaaaga	aaaattggag	agataaaaatt	agatttttgta	120
gttttttttt	ttttcgggaa	tgtttttttt	ttttttttta	gtttttgatg	aatgggttatt	180
atttattttt	attaaattta	aataaggatt	gttggtttgt	atgttttaatt	aggtaggttag	240
agggaaattgg	tttggttagg	aagtagtgat	tgagatggtt	tggttaagtt	agtgatagag	300
gaggggagaa	agaatttaga	ttaatttgta	tgtagtatat	tttattttta	tgaaataaaa	360
tatatattgtt	ttatatattgt	tgaaaagtaa	aataataata	ttgtacgaaa	tggtatataat	420
agggtaggtt	gtatatagta	gttttagaaa	tattattgta	tttatttagag	aaattatttt	480
aaaattgata	tttatataatt	ttttataata	ataataatat	gttagaaata	tatagtgtgg	540
tatttagtat	atatattttt	ttgttcgtta	gcgaaaaatt	ttaatcgttt	ttgtataata	600
tggtttattt	taaagtttaa	tttttaaaaa	tattgttgtg	atattattaa	taattgtttt	660
tataaaatta	atgtgatatt	togatatata	tatatatttt	tagttattta	aatgttaata	720
atgttaaat	taaaaaataa	taagtttata	gtaatgttaa	aatgttatat	ttagttaaat	780
atttgtttgt	gtatgtgttt	ttgtaattgt	tagaaatat	tgtagtgaaa	gatgttagat	840
attgaggata	tttttttgaa	attaaaggag	tttttttttt	gatttagtgg	tttttttttt	900
tttatatagt	tttttttttt	tttttttttt	ttagtgttta	cgatttttta	gtataatttt	960
tagtttttta	agggcggagt	tggtttattc	ggtaaggttt	taggatttctg	gcgttggtggg	1020
tcgggtttat	acgggtcggg	ttattgtata	ttggtaagta	tttaggttgg	aggtcggggtt	1080
ttgtacgttg	gcgtagtcga	agttggagtg	ttgttttggt	tttagtttta	ggttgggttag	1140
gttcgagtta	tacgtgtttt	tataaatata	cggaggagtc	ggcggcgcgt	aaggataggt	1200
agggcgtcgg	atcgcggaat	ttagcgaagg	gttatttagg	ttgttttaagt	tatttaggtt	1260
gttgagattg	gagttcggga	cgtttgttat	tggtgagggg	attatgtttg	acgatattgtt	1320
tatggacgag	atagagttag	gtggggaaaa	tatgttttgt	gatgataggg	ggttgacggt	1380
tatagagttg	aagaagggga	agtttttggt	ggatagggag	gcggatgtaa	ggtttttggc	1440
ggtttagttg	ttgtaggaat	agttttgggt	tatgtcgtcg	taggggttga	tgagtttatt	1500
gaattgcggt	tcgaagttat	ttttgtatag	ttcggtttgt	tggttgcggt	tttttttttt	1560
ttatttggtt	cgacgatttt	tgaattaaat	ttgggggcgg	ttggggtaag	ggagtaaaata	1620
gatgttatag	tgtagattat	taaaattttt	atcggaggtt	aattttcggt	ttttttcgat	1680
atatacgtta	gcgtattttat	atatttttgt	ttcgttttat	tgtatcggtt	tgtatattaa	1740
gatatttaggg	ttagttttta	gttattgggt	cgggttttta	ttaagcgtag	gagatttggt	1800
ttgttttggt	ttgcgagttg	ggattcggag	ttacgtttata	aattttagtc	gaacgtatgg	1860
agatttgcgg	acggtttgat	tatttagtta	ggcgtttttt	taggttttaa	aatatttaaat	1920
gtaaaataaaa	cgcggggtag	taggtttttt	taattttttt	cggggtattt	tgtaattttg	1980
tttttatttt	aaagtatatag	atttacggat	gaggagaagg	ggttggaagg	gtatttagagg	2040
atcgtttttt	tttttacgta	attttttttt	tttttttttg	atttttattg	tcgtttttta	2100

ttttttggta	cgtgtttttt	taatagggat	taggtcggtta	atattttttt	tcgttttagta	2160
aaataattaa	ataaagagta	aaagattatt	ttttcggttag	ttcggttaatt	ttaggagttt	2220
ggtatattaa	atttcgggaa	ttcggaaagg	gtagtttttg	agattttttt	ttttttcggt	2280
ttgttttttt	tttatttttaa	gtttattata	ggtttggttcg	cgcggttaggt	ttagtcgggt	2340
cgtttggttt	tgtaggcggt	tatttaggtc	ggtcgggtttt	tattcggtgtt	cgggtggttta	2400
gtcgtaatatt	cgattttta	ttatatcggt	ttttttgttc	gttttagacg	gcgggtttttg	2460
tgtattggag	agagggtttg	tttgagatat	tcgagttgat	attagtgatg	ttttatatta	2520
tatattttcg	tcgggttttag	tcgtgtaatt	cgtttttttt	tttttttttt	ttattttttga	2580
tttttttttt	attttttttt	ttttttgtat	tcgattgtta	taaaaagtac	gtttttatttt	2640
tatttggttc	gataagtagt	cgttttggaa	ggagaggtag	ttgtaaggag	agtttagcgt	2700
cgcggttata	aagtattagg	gtggagttgc	ggaatagcgg	gcggggtggg	agggcggttt	2760
cgaaggattt	tagaaaattt	atagattttg	tttttaatta	tttggtattt	ttatttttagg	2820
ttattttaaat	tttggttagg	cgagaagagt	acgtgagagg	ttcggttttt	tgatgtgtaa	2880
gagagttaat	gaaagattga	ttttgtttaa	aaattacgtc	tttaggattt	agttttggtt	2940
ttggatagtt	aaattaaaa	tattttta	tttttttcgg	ttttttattt	attagtagatg	3000
ttttatgttt	tgtataaatg	ttatttagag	agtgttttta	ttttttttga	tttgggagag	3060
tattttgggtt	tttatttttt	ttatcgttgt	tttttttttt	ttgtttgttt	tgttttaatc	3120
gggggtttta	ttttttttat	ttagagtatt	taattttttt	tttttaatag	ttaaagttttt	3180
ggatgtcggt	tgattttgtt	gattttgttt	ttgtttttta	gaattttaat	aaattttggaa	3240
ttttttatcg	attagtataa	attaggacgt	tggtattggg	ttattttattt	gagttttattt	3300
ttgttaattt	ataaagtata	gattttgttat	aaagttaagg	taagtttttt	ttataaaatt	3360
atgattataa	tttagaagag	ggggtgtgag	ttttaatttt	tagagtttta	tttttagagag	3420
aagataataa	aaataagtag	aaaagttttt	tttttttttt	tttttttttt	tttaagagga	3480
ttagtagttg	tgtattaaaa	ttttgttttc	ggagattata	aaattaggaa	ataggggtgtg	3540
tgggagagat	ttgaatggtc	gaaataatcg	taaagaaggt	gtaagaagcg	cgagtttagg	3600
agggaaaaag	ttgggttagg	gtcgggataa	agggttttta	gggaggggta	attttttcgt	3660
gttttttggcg	ggtttttttt	gttaaagggt	tataggttcg	agtttggttcg	cgggtttttgg	3720
tttggtaggg	attttattag	ttttgttttg	gtaattgtaa	gtaggaata	taatgttttg	3780
tgtaggggat	tgtttatgta	gttttagttcg	tgagatcgcg	ggatggcggg	gtagttagtc	3840
ggtgtcggtt	tgggagtttg	agttagggcg	gtagttttgt	cggtttcgga	gagggaaattg	3900
taatttcgta	attaggtcgt	cgcgaggttt	ttgttttttg	taaagttgcg	ttttatcggc	3960
gttttttttag	gcggcggttg	tttttatatt	tttttttggt	ttatttggtt	tgtattttta	4020
taatatattt	ttttattttt	tttttagattt	cgtgttggtt	tttttcgga	ttcgggtttt	4080
cgtaagggtg	gtttatatag	cgatttttttc	gctgtgtgat	atgttcgggt	agcgggtttt	4140
ttggaaagtg	gttttttagtt	tttgaggttg	ttgggttggt	aagttagttc	gttgcgtttt	4200
ttgtcggttt	tttttagacg	ggtttttcggc	gttttaacgtt	ttattttttt	tttgttggtt	4260
tttatttttt	tttgaaaacg	aaatatatat	attttttcgt	tagtatgttt	atttgtaacg	4320
cggacgttaa	ttggatcggc	ggtagaagtc	gtggaagagt	tgggttggtt	ggcggtcgag	4380
gaggggtgcgc	gcggcggttt	cgggtcgcga	ggagcggtgc	gtttgtgggg	tgtgtaggcg	4440
taagtgtggg	tgtttcgcgt	ttattttttt	ttttttttta	gcgtcgtagc	tttttatatt	4500
atgtttattt	tattgtagcg	gtatatattat	ttttatagtt	tgtgttttta	agtatatatta	4560
tatatttttg	cgtagatata	ttaaattttt	tgggacgcgt	atacgcgcgt	ggtttataga	4620
tttttttttt	tttcgtagaa	agtttagatt	tttatgcggt	ttgggaaggt	taggaaaaaga	4680
tgtggggatt	cgggttgggt	tcgaagttcg	tcggtttttt	tttaaaaaaa	aaaaaaaatt	4740
gttttttcgc	gaagggtatt	tttgagtggt	tttaggta	tttttaacga	gtggagtttt	4800
tcgggagttg	aaagtcgaga	ggaaaatagg	gatagaggtc	ggcggttttt	gaagggtttc	4860
gaattaagat	gttgggattt	ttgtgattta	ggaaatagaa	gggaggttag	ggtacgaata	4920
gagaggcgcg	tagaattgtt	cgcgttttta	cgcgttttagg	agtcgggtcg	gtcgaaggag	4980
aattaaaggg	atgcggggta	gttaaaattt	cggttttcgg	aagttttgcg	gggagtttag	5040
cgaacgatta	tttttattac	gttttttttc	ggaggggttg	attttttttg	ggcgagaggg	5100
agcgggtggc	gtagagtagt	tgagcgggaa	tgtttgtagg	gcggcgcggc	gtttttatttg	5160
cggttttcgg	gttggagggtg	tcggagatgg	tgtgtatttt	tagtttgtgt	ttggaggagt	5220
ttagcgatcg	gggttgatcg	ggagttagaa	tcgaagttat	ggtttaacggt	ttggggtggt	5280
gataggaaga	tgaggagacg	gtcgatagtt	tgggttttcgt	tgttcgggtg	tttaagtga	5340
gcgggttttt	tatgtagttt	atggacgagg	gagcgcgacg	ttttattagt	ttttggttat	5400
tgtttcgtcg	agttttcgta	gtcgtcggtg	ttcgttttcg	gtcgcgtttt	agggcgcgag	5460
ttttttcgtt	gcggggagag	ttaggggacg	taattttcgt	cgagttttta	agttaaagt	5520
ttttcgtttt	tttcgggaagg	tttaagcgaa	aaagttcgga	gacggaaagt	tagcgggtta	5580
acgaagatat	gggatgtggg	tagaagggt	ttatttagag	cgttttttag	gagtaggttt	5640
tttaagtttta	aagcgaaata	agagtggtta	aagatttttt	tttttttttt	tttttttttt	5700
taagaatttt	tttaataagg	aaagtttaac	tcgatcgcg	tttggttcgt	ttttttttac	5760
gcggtagttt	tgatagagaa	gtgttaagag	tgatagggat	aggtaggtga	tattagattt	5820
tttgcgcgcg	tagtagtcgt	tgtagttacg	acgcgggttt	ttgagcgtat	ttttcgtaac	5880
gcgtatacgt	atattttttc	ggcggtcgaa	taggagtcgg	gttttgctgt	agtttagttt	5940
taggtatttt	ggcgagcgac	ggatttagatt	tcgggttttc	cgtttttttg	ttggttta	6000
atttttaaaa	tagaggcggt	ttttttggtg	tcgagacgtt	atttcgtcgc	gttttttttt	6060
agtttttttc	gttttcgttt	tttttttagat	tttttttcgg	gtgcgattga	cgtgggttcg	6120
tattaattag	gacgtttcga	gtcgcgggtg	agggattgtt	ttgtttgtat	ttatttagtag	6180
tgccgggtcg	ggttattgtt	tcgtcgtgcg	tattgggttt	atataggtaa	gttttcggga	6240
attttagttt	tgttttagtt	aaggcgattc	ggttttttag	aegaatttaa	aggtgaagag	6300
atgagggttag	gagtcgaagg	tttgggagaa	gagagtggaa	tggttaagaa	gagaaaggta	6360

taaggattaa	taagatattt	atTTTTtGtG	ttttattata	tttatttttta	atTTTTtatt	6420
ttatataaaa	aggagatacg	ttattttaaaa	ttagaaaatt	tgaaaaatag	taataaatta	6480
tttttttogat	tttaaatttt	ttaaatagtt	tgtaaagtga	atgttgcggt	aatttgaaga	6540
agttttaatt	gtaaagaaga	tagagttttg	aaaaggtagg	ttaatataatt	agaaatcgag	6600
aagtaaattgg	attcgttaaa	agaaaattat	tttgatttta	aacgaataat	tgtttgggtgg	6660
tttatttttgg	atTTtatataa	gaataaaaag	tCGtttttaga	ttacgttttt	tgtgatgttt	6720
attagttttt	agatagaaaa	tatataatag	aagagaaatt	ttaatTTtagc	gttttttaaaa	6780
tgttgaaagt	ttattttattt	tattttaacgt	tgattaaagat	atatattttta	gatttttttaa	6840
atTTTTtGtG	tattgtatta	agttcgtttt	aattcGagag	agttacgttt	taaattcGat	6900
tttttttGttt	atTTTTtattt	taattagatt	taaattttata	aagtttGtag	aattaataat	6960
tttgagtttaa	ttatatatga	aatatgtttt	aatgaatttt	tatataatta	agaatgtttgt	7020
taaataatta	atTTttaagga	taattttttaa	tagttatttt	tttttttttag	tgagtttaag	7080
gttgtttttga	gttatttaaag	tttaagtagg	tagaaggggt	gtgtgtgagt	taagggcgaa	7140
aagtttagaa	ttgcgtttta	ttagtataag	taaaattttta	tttatataaaa	ataaaaaaaa	7200
ttatttttgg	agataattaat	tttttatagt	attgtttttta	agtaaaattta	atTTtttaaag	7260
aaattaaaga	aagaaatttta	aatatatttta	aaataattttt	tgaaagttttt	tttGttttttt	7320
agtatagggt	agttggagag	gataaatttaa	tttttttttgg	gttttttGtat	gggGgattgt	7380
tttatttatgg	agttagtgtt	attattttttg	aatgtgtattt	tgtttGtatat	tatagttaat	7440
gatttGtaaat	gttagtatga	agtatttttta	aaatatTTttt	ttttttGtttt	tgttttataag	7500
attgggaaat	ttattcGatg	tggaataaag	tggatgaagt	agattataaaa	tatatTTtGta	7560
atTTtatgtgt	tttttttttGt	ttttgatttat	tttttaaattt	tatttGtaaat	tttttttttat	7620
tttaaattttg	tagttttaaaag	acgtatatga	gaattgttttt	ttagttttttt	tttatttagta	7680
ttatttttatt	ttagaataaa	tttagttGta	agggaggaat	tttttttatag	taagtttttaa	7740
attagtatttt	ttgttttttaa	tttttttattt	tatttttattt	tatttttatat	atatagatat	7800
ttgttttagag	taaaatatat	ttttatgtga	taggtttGta	ttagttgagg	tttatatatt	7860
tagtttatatt	aggtttttGta	atTTTTtattt	taaattatat	atatttatatt	agtagtttGt	7920
tggtaaagaa	ggttaaattta	atTTtatattt	tgttttatttat	ttgggtgttta	aatgacgtat	7980
tttattttcgg	agattttggcg	gagaatttttt	tttttagattt	ttatagcgtt	ttattgaaga	8040
taatgtttttt	atattttGtag	tggttttttaa	tttgataaga	ttttaattttg	tttaagttttt	8100
ttaaataagg	gttttaaatg	tttttagtcg	ttttttttatt	gaatttttttt	taatttttttt	8160
aagattataaa	agtatatgtg	taaagtaaat	atTTTTtTTtt	attgtatttgt	tagtcGatga	8220
tttataatta	agttataaag	aattttagttt	tttttttGttg	aatgtgtttta	ttaatatat	8280
tttagtttttt	tttttttaaat	tttagaataag	ttgtggttttt	tataatatata	tgtttttttaa	8340
agtttttattt	tatgaagggga	ttttattata	ttaaagaaatg	aaaaaaatttt	ttattGtagt	8400
tagttatatat	agtttttttat	tttttGtttt	ttaaagattta	aatttttagag	ttgtaaatat	8460
ttttgggaagt	ttgggtgttta	atgtttttatt	ttagaaagtc	gagaagttttt	atagagttat	8520
atagatttttt	aaattttattt	tttataaatt	tatagaattt	tgataaaaagt	tttggtgggtt	8580
ttatttttatc	gatggaaattt	ttattacgat	aaatatatat	gtatgaaggga	ttttaatttag	8640
ttttttaaagt	ggttgaaaaa	tttaagggta	cgtgattgtt	tttttatagt	ttaacgtgtg	8700
cgagatgttg	gaagtattgg	ggattagtag	tagtttagat	gttttaaaaag	ataaggtgtt	8760
ttaatTTtGtG	tggattttatt	gaagttaagt	ggtgaataaaa	gataattatt	tagataattt	8820
agattaaagt	aaaagtataaa	ttatatattat	ttgtatatat	atattttatat	tttttttata	8880
ttatagatat	atatacgtat	atatatattg	gtttttGtaaa	taattgattt	aaagtgagga	8940
tttttttttGt	atTTTTttag	taggagtttt	aatattttttt	taattttttta	attatttttat	9000
a						9001

<210> 3

<211> 9001

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 3

tgtaaagtga	ttaggagatt	aggagaatgt	tgaaatttttt	gttggaaaaa	tgtaaagaaa	60
atTTTTtattt	tgagtttagtt	gttttatagag	ttagtgtgtg	tgtgcgtgtg	tgtgtttGta	120
atataaaaatg	gatgtgaata	tatatatata	aatagatatg	gttttGtttt	tatttttaatt	180
tgaaattattt	agataattgt	ttttattttat	tatttgattt	taatgggttt	atataaatta	240
ggatatTTtta	ttttttttagg	tatttagggt	gttggttgatt	tttagtgttt	ttaatatttc	300
gtatacgttg	gtattatgag	gagtagttac	gtgttttttgg	gtttttttaat	tattttggag	360
gttgattgag	gtttttttata	tatgtatatt	tgtcgtgatg	aaagttttat	cggtagagtg	420
gagttatttag	agttttttatt	aaaattttgt	gggttttatga	gagatgggtt	tagaaaattg	480
tatgggttttt	tgggggttttt	cgggtttttta	aaataaggta	ttaatatttta	agttttttaaa	540
aatattttGta	gtttttggggg	ttgaatttttG	aaaaataagg	agtgaggggt	tgtgtatat	600
aatttatagtg	gagattttttt	ttattttttta	atgtgatgga	gttttttttat	gaaatgaagt	660
tttaaggggtg	atggattttgt	ggggattata	gttatttttga	ggtttaaaaag	aagaaattgg	720
aatatgatta	gtaaaatatat	ttagtagaaa	agagttggat	ttttattgat	ttagttatag	780
gttatcgggt	ggtagtgtaa	tgggaggaaa	tattttattt	atatatatat	tttatgtatt	840
tgggggaatt	agaggaaatt	taataagaaa	acggttagaa	atattttaaaa	tttttatttta	900

aaagatttta	gtaaattaga	gttttattag	attaaaaatt	attataaatg	taagagtatt	960
gttttttagt	aaacgttgtg	gggttttgaga	aggagatttt	tcgttaaatt	ttcgggataa	1020
aatgcgttat	ttaagtatta	gataatgagt	agaatgtaaa	ttaatttaatt	tttttttatt	1080
aataggttgt	tagtgtaaatg	tgtataatttt	agtataaaga	ttgtaggatt	taatatagtt	1140
ggatgtatga	gttttagtta	atgtagatttt	gttatatgag	gatgtgtttt	attttgagta	1200
gggtgtttgta	tgtgtggaat	gggtgaaagt	ggaataaaag	gttaaaagta	gaaatgttga	1260
tttaaagttt	attatgaaga	aattttttttt	ttgtagttaa	attatttttta	aagtgggatg	1320
atattgggtga	agaaagattg	aaaaataaatt	tttatgtgctg	tttttggatt	gtaagtttaa	1380
aatggggagg	agttgttagat	agggtttggg	gggtggttagg	gtaaaggaga	gatatataag	1440
ttgtaaatat	atttgtagtt	tgttttatttt	attttgtttt	atatacgaata	agtttttttaa	1500
ttttgtgaat	aaggataagg	agggtgtgtt	ttaaagatat	tttatgtttg	tattgtaaat	1560
tattgattgt	aatgttaaatt	aaatatatat	ttagagatga	taatattaat	tttatagtaa	1620
aataatcggt	tatgtagaaa	tttagaggag	attagtttgt	tttttttagt	tgatttatgt	1680
tgggggataa	aaggattttt	aaaaattatt	ttgaatatgt	ttggattttt	tttttttatt	1740
tttttggaaa	ttaaatttgt	ttggaaatag	tgtataaaag	agttgatgtt	tttaaagggtg	1800
atttttttttg	ttttatataa	ataaggtttt	gtttttgtta	gttgagcgtg	gttttaggtt	1860
tttcgtttttt	agtttatata	tattttttttt	gtttgttttg	atttttaattg	tttaagatag	1920
ttttgagttt	attgggaaaa	gaaaatgatt	gttaaaaaatt	atttttgaaa	ttggttattt	1980
ggtaatatatt	ttaattgtat	ggaaatttat	taaggtatat	tttatatata	attagtttaa	2040
gggtgttgat	tttatagggt	ttatggattt	aaatttgatt	gataataaag	taaataagag	2100
agtcgaattt	aaagcgtggt	tttttcgggt	taggacgagt	ttaatatagt	gtataaggaa	2160
tttgaaagat	ttaggatatg	tgttttaatt	aacgttaagt	agaatggata	agtttttagt	2220
attttgaaaa	cgttgggtta	gggtttttttt	tttatgtgtg	gtttttgttt	tggggattaa	2280
taagtattat	agagaacgtg	atttgaggcg	attttttatt	tttgtataaa	tttagagtga	2340
attattaaat	agttgttcgt	ttaaagttaa	ggtaattttt	ttttgacggg	tttatttgtt	2400
tttcogatttt	taatttatta	gtttgttttt	ttagggtttt	gttttttttg	taattaaagt	2460
tttttttagat	tagcgtagta	tttatttgat	aggttgtttg	gaaaatttaa	gatcggagag	2520
gtgatttgtt	gttgtttttt	aaattttttt	gtttttaagta	acgtgttttt	tttttatatg	2580
gggtggggga	ttggaaatgg	atgtagttag	atataaaagag	tgggtgtttt	gttgattttt	2640
gtatttttttt	ttttttgatt	attttatattt	tttttttttaa	gttttcgatt	tttagtttta	2700
tttttttatt	tttgggttcg	tattaaaaagt	cggatcgttt	tgggttgggt	aggagttgaa	2760
tttttcgggag	tttgtttgtg	tagattttagt	gcgtacggcg	aggtagttagt	tcggtttcgt	2820
attgttgata	ggtgtaggta	ggatagtttt	tttatcgcgg	ttcggggcgt	tttgattggt	2880
gcggagttac	gttagtcgta	ttcggagaag	ggtttgggag	gaggcggagg	cggagagggt	2940
tggggagggt	cgcggcggag	tgacgtttcg	gtattaggaa	gttcgttttt	ggttttaaga	3000
tgttaggtta	atagggaagc	gcggagtcgt	agatttgggt	cgtcgttcgt	ttgggtgttt	3060
ggagttgagt	tgcgtaagg	ttcggttttt	gttcgatcgt	tcgaggggtg	tgcgtgtcgc	3120
cgttgccggag	ggtgcgttta	gagggtcgcg	tcgtggttgt	agcggtttgt	gtcgtcgtag	3180
gggatttaatt	attattttatt	tgttttttgt	attttttgata	tttttttgtt	agggttgtcg	3240
cgtggggggg	gggcgggtag	agcgcggtcg	gcgttagttt	tttttatttg	aggggttttt	3300
gggggaggga	gggagagaag	aaggggggtt	ttgtttattt	ttgtttcgtt	ttggagtgtg	3360
gaagtttgtt	ttttaagac	gttttgagtg	gtgttttttt	gttttatatt	tatgttttcg	3420
tttgttcgtt	gatttttcgt	tttcggattt	tttcogtttg	gttttttcgga	ggagacgggg	3480
gtagtttggt	ttgagaattc	ggcgggggtt	gcgttttttg	gtttttttcgc	tagcggggaa	3540
attttcgcgtt	tagagcgcga	ttcggagcgg	gtagcggcgg	ttacgggggt	tcggcggggg	3600
agtagttaag	gattagtaga	gcgtcgcgtt	ttttcgttta	tgaattgtat	gaaagggttcg	3660
ttttattttgg	agtatcgagt	agcggggatt	aagttgtcgg	tcgtttttttt	atttttttgt	3720
tattatttttt	agtcgttagt	tatggtttcg	gtttttgggtt	tcggttagtt	tcggtcgttg	3780
gatttttttta	agtatagggt	ggaggtgtat	attatttttcg	atattttttag	ttcggagggtc	3840
gtaggtaagg	cgtcgcgtcg	ttttgtagat	atttttcgtt	agttgttttg	cgttatcgtt	3900
tttttttcgt	tttaaggaa	ttagtttttt	cggggggagg	cgtgggtggga	gtggtcgttc	3960
gttttggtttt	tcgtagaatt	ttcggggagtc	gggaattttga	ttattttcgt	tttttttagt	4020
tttttttcga	tcggttcggt	ttttggggcg	ttaaagggcgc	gagtaatttt	gtcgtttttt	4080
ttatttcgtat	tttgggtttt	tttttgtttt	ttgggttata	aaaatttttag	tattttgatt	4140
cgaggattttt	tagaggtcgt	cgattttttgt	ttttgttttt	ttttcgggtt	ttagtttttcg	4200
aggagttttta	ttcgttagga	aattgttttg	aattattttag	aaatgttttt	cgcgaaggag	4260
tattttttttt	tttttttttg	gaaaggggtc	gcgaattttcg	gtgtttaatc	gaatttttat	4320
atttttttttt	agtttttttta	aatcgtatgg	aaattttgagt	tttttgcgag	ggggaggggg	4380
gtttgtaaat	tacgcgcgtg	tgcgcgtttt	aggagatttg	gtgtgttttc	gtagagggtg	4440
ataaatatat	ttgaaagtat	aggttataaa	agtgaatgtg	tcgttgtagt	gagataaata	4500
tgtaaataaa	acgtgcggcg	ttggggggag	ggaggaaatg	gggcgcggat	atttatatatt	4560
gcgttttgat	attttatagg	cgtagcgttt	ttcgcgggtc	ggagtcgtcg	cgcgtatttt	4620
tttttcggcgt	taggtagttt	agtttttttta	cggtttttgt	cgtcgggtta	gttggcgttc	4680
gcgttgtagg	tgggtatggt	gacgggaaag	tgtgtgtgtt	tcgttttttag	agaaagataa	4740
aagtttagtag	gggaagaatg	aggacgtggg	cgtcagggat	tcgtttaaga	agaagcggta	4800
aaggcggtag	cggattttatt	ttattagtta	gtagtttttag	gagttggagg	ttatttttta	4860
gagggaatcgt	tattcggata	gttttatacg	cgaagaaatc	gttgtgtgga	ttaatttttac	4920
ggaagtccga	gttcgggtag	gagttagtag	ggagtttggg	agggatgggg	ggaggatggt	4980
gtggagggtat	aggttaagta	gattaggaga	gaatgtggaa	ggtagcgtcg	tttggggagg	5040
cgtcgggtggg	gcgtagtttt	gtaaaggtag	aaggtttcgc	ggcggtttg	ttgcgagatt	5100
atagttttttt	tttcgaggtc	gataggattg	tcgtttttggt	ttaggtttttt	agagcgggat	5160

cgggtttattg	tttcggttatt	tgcgcgatttt	acgaggttggg	ttgtatgggt	aattttttgt	5220
ataggatatt	gtgttttttgg	tttgtagttg	ttagagtaga	gttaataaaa	tttttattag	5280
gttaagagtc	gcgaataggt	tttaatttgt	gagtttttaa	taaggaaaaa	tcggttagaga	5340
tacggaagag	ttgggtttttt	ttgggaaatt	tttgtttcgg	ttttgggtta	gttttttttt	5400
ttttgggttc	gcggtttttta	tatttttttt	acggtttgtt	cgggttattta	gggttttttt	5460
atatatttta	tttttttagtt	ttgtgatttt	cgggagtaaa	gttttaatat	ataattatta	5520
gttttttttag	aaggagaaaag	aaaaaaaagaa	gaaagatttt	tttgttttgg	taattttatt	5580
tttttttagga	gttgaattttt	ggaaattgaa	atttatattt	tttttttttaa	attataatta	5640
tagtttttgta	aaaagggtttt	attttaatttt	tgtagtaaat	ttgtattttta	tggattggta	5700
aaaatgagtt	taaataaata	atttaatagt	aacgttttgg	tttatgttgg	tcggtggaag	5760
atttttaaatt	tgttaggatt	ttggaagtag	aaaatagaat	taagtaaatt	aagcgggtatt	5820
tagaggtttt	gttgttaaaa	aaaaaaaatt	aagtgttttg	ggtagaaaaa	ataaagtttt	5880
cgggttagagt	agagtaaaata	aaaagaagaa	aataacgata	aaaagaataa	agattaaaat	5940
gttttttttaa	atttagaggga	atgaagatat	tttttgggtg	gtatttgtgt	aagggtatgag	6000
gttatgttgg	tggataaaaag	gtcgggaaga	agttgaaaaa	gggttttagtt	taatttttta	6060
gagtttaggg	tggttttttgg	gcggcggtgt	tttgagtaag	gttagttttt	tattagtttt	6120
tttgtatatt	aagggaacgg	gtttttttacg	tattttttttc	gtttgagtaa	agtttagatg	6180
gttttagggta	gaaatggtaa	gtaattaaag	atagagttta	tgggtttttt	gggatttttc	6240
gaaaacgctt	ttttattttcg	ttcgttattt	cgtagttttta	tttttagtgt	tttgtagtcgc	6300
ggcggttgggt	ttttttttgta	gttgtttttt	tttttagggc	gggtgtttgt	cgagtttaagt	6360
gggagtgagg	cgtgtttttt	atagtagtcg	gggtgtaaaga	ggaaggggga	taaaaaggaa	6420
attaagaatg	aaaggaaaaa	gagaaaaagc	ggattatacg	gttgggttcg	gcggagatgt	6480
gtaattgtgaa	atattatttg	tgtagtttcg	gatatttttag	gttaggtttt	tttttaatat	6540
ataaaagtcg	tcgttttggcg	cgataggggg	gttcgatgtg	gattgggatc	ggggttcggg	6600
ttgggtttatc	ggggttttgc	ggaagtcggg	cgggtttgggt	gggtcgtttgt	aaaggttaaac	6660
gattcgggttg	gggtttggcgc	gcggataggt	ttgtggttgg	tttagggtaa	agaagaggta	6720
gagcgaaaga	aggggggaatt	tttaaaatta	tttttttcgg	gttttcggag	tttaatatgt	6780
taagttttttg	gagttaacga	gttgacgaag	aggtggtttt	ttgtttttta	ttgggttgtt	6840
ttgttaggcg	agaaagagtg	ttggcggttt	agttttttgt	aagggtgtac	gtattagggg	6900
gtgggggacg	atagtggagg	ttagggaagg	aagggtaggaa	ttgcgtggga	gaaagagcga	6960
tttttttagtg	tttttttagt	tttttttttt	tattcgtggg	tttgtggttt	tggaaatggaa	7020
gtaagtttgt	aaggtgtttc	gggaagggtt	ggaaaagttt	gttgtttcgc	gtttgtttta	7080
tattaaagtgt	ttttggattt	ggagaaacgt	ttggttgagt	gattaaatcg	ttcgtagggt	7140
tttatgctgt	cgggttgaggt	ttgtggcgta	gtttcagagt	ttagtctgta	gggttagagta	7200
gattaggttt	tttgcgtttg	gtggagattc	gggttagtaa	ttgaaagtgt	gttttggtat	7260
tttgggtgtgt	agggcggtgt	agtgaagcga	gggttaggggt	tgtgagtgcg	ttagcgtgtg	7320
tgtcggggga	agggcggggt	tgggttttcga	tggaaagtgt	agtaatttgt	attgttggtat	7380
ttgtttgttt	tttgttttta	atcgttttta	gggttgggtt	aagaatcgtc	gggttaaatg	7440
gagaaagagg	gagcgttaatt	agtaggtcga	gttatgtaag	aatgggtttc	gggtcgtagtt	7500
taatgggttt	atgtagtttt	acgacgatatt	gtattttagg	tattttttata	ataattgggt	7560
cggttaagggt	tttatattcg	ttttttttatt	tattaagagt	tttttttttt	ttaattttat	7620
gaacgttaaat	tttttgttat	tatagagtat	gtttttttta	tttaattttta	tttcgtttat	7680
gagtatgtcg	tttagtatgg	tgttttttagt	agtgataggc	gtttcgggtt	ttagttttta	7740
tagtttgaat	aatttgaata	atttgagtag	ttcgtcgttg	aattttcgcg	tgtcgacgtt	7800
tgtttgtttt	tacgcgtcgt	cgatttttttc	gtatgtttat	agggatacgt	gtaattcgag	7860
tttggttagt	ttgagattga	aagtaaaagta	gtatttttagt	ttcggttacg	ttagcgtgta	7920
gaattcgggt	tttaatttga	gtgttttgta	gtatgtagtg	gatcgggttcg	tgtgagtcgt	7980
atztatagcg	tcgggatttt	aggatttttgt	cggatgggggt	aattttcgtt	ttgaaagatt	8040
gggaatttat	ttagaagggtc	gtgggtatta	aagaaaggga	gagaaagaga	agttatatag	8100
agaaaaggaa	attattgaaat	taaagagaga	gttttttttga	tttttaaagg	atgttttttag	8160
tgtttgatat	tttttatttat	aagtattttt	aatagttgta	aggatatata	tataaataaa	8220
tgtttgattg	gatatgatatt	tttaatatata	ttataagttt	gttattttttt	aagtttagta	8280
ttgttaatat	ttaaatgatt	gaaaggatgt	atatatatcgt	aaatgttaaa	ttaatttttat	8340
aaaagtagtt	gttagtaata	ttataatagt	gttttttaaa	gttaggtttt	aaaataaagt	8400
atgttatata	gaagcgatta	ggattttttcg	tttgcgagta	agggagtgtg	tataattaaat	8460
gttatattgt	atgttttttaa	tatatatttta	ttattataaa	aaatgtgtga	atattagttt	8520
tagaatagtt	tttttgggtg	atgtaatgat	gttttttgaaa	ttgttatgta	taattttatt	8580
tgtgtataat	atttcgtata	atattatttgt	tttatattttt	agtaaatatg	aaataaatgt	8640
gtttttatttt	atgggagtaa	aatatatattg	atataaaattg	gttttggtttt	tttttttttt	8700
tttttgttat	taatttgggt	aggatatattt	agttatttgt	ttttaaataa	attagttttt	8760
tttgtttgtt	tagttaaata	tataaggtag	tagttttttat	ttaaatttgg	tagaaataaa	8820
tgtagtttat	ttattagaaa	ttaaaaagaa	aaaaaaagggt	attttcgggg	gggaaaagggt	8880
ttataaaaatt	taatttttgt	tttttaatttt	ttttttgggt	taaattttaga	ggattttatt	8940
atggtttagta	aataatatga	aaaagaaaaa	agaagaaaga	aatttagtaa	gtttattagt	9000
t						9001

<210> 4
 <211> 9001
 <212> DNA
 <213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 4

agttgatgga	tttggttaaat	ttttttttttt	ttttttttttt	tttatattat	ttgttagtta	60
taatggaatt	tttttaggttt	aagtttaaaga	aaaattggag	agataaaaatt	agatttttgta	120
gtttttttttt	tttttgggaa	tgttttttttt	tttttttttta	gttttttgatg	aatgggttatt	180
atttatttttt	attaaatttta	aataaggatt	gttggttttgt	atgttttaatt	aggtaggtag	240
aggggaattgg	tttggttagg	aagtagtgat	tgagatgttt	tggttaagtt	agtgatagag	300
gaggggagaa	agaatttaga	ttaatttgta	tgtagtatat	tttatttttta	tgaaataaaa	360
tatatattgtt	ttatatattgt	tgaaaagtaa	aataataata	ttgtatgaaa	tggtatataat	420
agggtaggtt	gtatatagta	gttttagaaa	tattattgta	tttatttagag	aaattattttt	480
aaaatttgata	tttatatatt	ttttataata	ataataatat	gttagaaata	tatagtgtgg	540
tatttagtat	atataattttt	ttgtttgtaa	gtgaaaaatt	ttaattgttt	ttgtataata	600
tggttttattt	taaagttaa	tttttaaaaa	tattgttgtg	atattatttaa	taattgttttt	660
tataaaatta	atttgatatt	ttgatataata	tatatattttt	tagttatttta	aatgttaata	720
atgtttaaatt	taaaaaataa	taagtttata	gtaatgttaa	aatgttatata	ttagttaaatt	780
atgtgtttgt	gtatgtgttt	ttgtaattgt	tagaaatatt	tgtagtgaaa	gatgttagat	840
attgaggata	ttttttttgaa	attaaaggag	ttttttttttt	gatttagtggt	ttttttttttt	900
tttatatagt	ttttttttttt	ttttttttttt	ttagtgttta	tgattttttta	gtataattttt	960
tagttttttta	aggggtggagt	tggttttattt	ggtaagggttt	taggatttttg	gtgtgtgtggg	1020
tggtgttttat	gtgggttggt	tatttgata	ttggtaagta	tttaggttgg	aggttgggtt	1080
ttgtatgttg	gtgtagttga	agttggagtg	ttgttttgtt	tttagttttta	gggtgtgttag	1140
gtttgagtta	tatgtgtttt	tataaatata	tgaggaggtt	gggtgtgtgt	aaggataggt	1200
aggtgttgtt	attgtggaat	ttagtgtatg	gttatttagg	ttgttttaagt	tatttaggtt	1260
gttgagattg	gagttttggga	tggtttgttat	tggtgaggtt	attatgttgg	tatgatgttt	1320
tatggatgag	atagagttgg	gtggggaaaa	tatgttttgt	gatgataggg	gggtgatgtt	1380
tatagagttg	aagaaggggga	agtttttggt	ggatagggag	gtggatgtaa	ggtttttggt	1440
ggtttagttg	ttgtaggaat	agttttgggt	tatgtttgtg	taggggttga	tgagtttatt	1500
gaattgtggt	ttgaagttaa	ttttgtatag	tttggtttgt	tggtttgtgt	ttttttttttt	1560
ttattttggt	tgtatgtttt	tgaattaaat	ttgggggtgg	ttggggtaag	ggagttaaata	1620
gatgttatag	tgtagattat	taaaattttt	attggaggtt	aattttttgtt	ttttttttgat	1680
atatatgtta	gtgtatttat	atatttttggt	tttgttttat	tgtattgttt	tgtatatttaa	1740
gatattaggg	ttagttttta	gttatttggtt	tggttttttta	ttaagtgtag	gagatttggt	1800
ttgttttggt	ttgtgagttg	ggattttggag	ttatgttata	aatttttagtt	gaatgtatgg	1860
agattttgtg	atgggttgat	tatttagtta	ggtgtttttt	taggttttaa	aatattttaat	1920
gtaaaaataaa	tgtggggtag	taggtttttt	taattttttt	tgggggtattt	tgtaaaatttg	1980
ttttttatttt	aaagttatag	atttatggat	gaggagaagg	ggttggaagg	gtatttagagg	2040
attgttttttt	tttttatgta	attttttttt	ttttttttttg	attttttattg	ttgttttttta	2100
tttttttggt	tgtgtttttt	taatagggat	taggtttgta	atattttttt	ttgttttagta	2160
aaataatttaa	ataaagagta	aaagattatt	tttttgtag	tttggttaatt	ttaggagttt	2220
ggtatatatt	attttgggaa	tttggaaagg	gtagtttttg	agattttttt	tttttttggt	2280
ttgttttggt	tttatttttaa	gtttattata	gggtttgtttg	tgtgttaggt	ttagtgtgggt	2340
tggtttgggt	tgtaggtggt	tatttaggtt	gggtgggttt	tatttgtgtt	tggtgttttta	2400
gttgtaaatt	tgatttttaat	ttatattggg	tttttttgtt	gttttagatg	gtgggtttttg	2460
tgtattggag	agaggttttg	tttgagatat	ttgagttgat	attagtgatg	ttttatatta	2520
tatatattttg	ttgggttttag	ttgtgttaatt	tggttttttt	ttttttttttt	ttattttttga	2580
ttttttttttt	atttttttttt	ttttttgtat	ttgattgtta	taaaaagtat	gtttttatttt	2640
tattttggttt	gataagtagt	tggttttgga	ggagaggttag	ttgtaaggag	agtttagtgt	2700
tgtggttata	aagtattagg	gtggagttgt	ggaatagtgg	gtgggggtggg	aggggtgtttt	2760
tgaaggattt	tagaaaattt	atagattttg	tttttaatta	tttggtattt	ttatttttagg	2820
ttattttaaat	tttgttttag	tgagaagagt	atgtgagagg	tttggtttttt	tgatgtgtaa	2880
gagagttaat	gaaagattga	ttttgttttaa	aattatgttg	tttaggattt	agtttttggtt	2940
ttggatagtt	aaattaaaaat	tattttttaat	tttttttttg	tttttttattt	attagtatag	3000
ttttatgttt	tgtataaatg	ttatttttag	agtgtttttta	tttttttttga	tttgggagag	3060
tatttttggtt	tttattttttt	ttattgttgt	ttttttttttt	ttgttttggtt	tggttttaatt	3120
gggggttttta	tttttttttat	tttaggtatt	taatttttttt	ttttttaatag	taaagtttttt	3180
ggatgttgtt	tgattttgttt	gattttgtttt	tttgtttttta	gaatttttaat	aaattttggaa	3240
tttttttattg	attagtataa	attaggatgt	tggtattggg	ttattttattt	gagttttattt	3300
ttgttaattt	ataaagtata	gattttgttat	aaagtttaagg	taagttttttt	ttataaaaatt	3360
atgattataa	tttgaagag	ggggtgtgag	tttttaatttt	tagagtttaa	tttttgagag	3420
aagataaata	aattaagtag	aaaagttttt	ttttttttttt	ttttttttttt	tttaagaggga	3480
ttagtagttg	tgtattaaaa	ttttgttttt	ggagattata	aaattaggaa	ataggggtgtg	3540
tgggagagat	ttgaatgggt	gaaataattg	taaagaaggt	gtaagaagtg	tgagtttagg	3600
agggaaaaaag	ttgggttagg	gttgggataa	aggttttttta	gggagggtta	attttttttgt	3660
gttttttggtg	gggtttttttt	gttaaagggt	tatagggtgg	agttttgtttg	tggttttttgg	3720
tttggtaggg	atttttattag	ttttgtttttg	gtaattgttaa	gttaggaata	taatgtttttg	3780
tgtaggggat	tgtttatgta	gttttagtttg	tgagattgtg	ggatgggtggg	gtagttagtt	3840
ggtgtttgttt	tgggagtttg	agttagggtg	gtagttttgt	tgggttttggg	gaggggaattg	3900
taatttttgta	attaggttgt	tgtgaggttt	tttgtttttg	taaagtttgt	ttttatttggt	3960

gttttttttag	gtggtgttgt	tttttatatt	tttttttggg	ttatttgggt	tgtattttta	4020
taatatTTTT	tttttatttt	tttttagatt	tgtgttgggt	tttatttggg	tttgggtttt	4080
tgtgaagggtg	gttttatatag	tgattttttt	gtgtgtggat	atgtttgggt	agtgggtttt	4140
ttggaaagtgt	gttttttagtt	tttggagttg	ttgggtggta	aaagtgaagtt	gttgtgtttt	4200
ttgttgtttt	tttttttagatg	ggttttttt	gtttatgttt	ttattttttt	tttgttgggt	4260
tttattttttt	tttgaaaatg	aaatatatat	atttttttgt	tagtatgttt	atttgtaatg	4320
tggatgttaa	ttggattgggt	ggtagaagtt	gtggaagagt	tgggttgttt	ggtgttggag	4380
gaggggtgtgt	gtggtgggtt	tgggttgtga	ggagtgttgt	gtttgtgggg	tgtgtagggt	4440
taagtgtggg	tgtttgttgt	ttattttttt	ttttttttta	gtgttgtatg	ttttatttat	4500
atgtttattt	tattgtagtgt	gtatatttat	ttttatagtt	tgtgttttta	agtatattta	4560
tatatTTTTg	tgtagatata	ttaaaatttt	tgggatgtgt	atatgttgtgt	ggtttataga	4620
ttttttttttt	ttttgtagaa	agtttagatt	tttatgttgt	ttgggaaggt	taggaaaaga	4680
tgtgggggatt	tggttgggta	ttgaagtttt	ttgggttttt	tttaaaaaaa	aaaaaaaatt	4740
gttttttttgt	gaagggtatt	tttgagtggg	tttaggtaat	tttttaataga	gtggagtttt	4800
ttggggagttg	aaagttgaga	ggaaaatagg	gatagagggt	ggtgggtttt	gaagggtttt	4860
gaattaagat	gttgggattt	ttgtgattta	ggaaaatagaa	gggagggttag	ggtatgaata	4920
gagagggtgg	tagaattgtt	tgtgttttta	gtgttttagg	agttgggttg	ggtgagggtg	4980
aattaaaggg	atgtggggta	gttaaaaatt	tgggtttttg	aagtttttgt	gggaggtagg	5040
tgaatgatta	tttttattat	gttttttttt	ggagggttgt	attttttttg	gggtagaggg	5100
agtgggttgt	gtagagtgt	tgagtgggaa	tgtttgtagg	gtgggtgttgt	gttttatttg	5160
tgggttttttg	gttggagggt	ttggagatgg	tgtgtatttt	tagtttgtgt	ttggaggagt	5220
ttagtgtttg	gggttgattg	ggagttagaa	ttgaagttat	ggttaatgtt	tgggaggtgt	5280
gatagggaaga	tgaggagatg	gttgatagtt	tgggttttgt	tgtttgttgt	tttaagttaa	5340
gtgggtttttt	tatgtagtgt	atggatgagg	gagtgtgatg	ttttattagt	ttttgtttat	5400
tgttttgttg	agtttttgtg	gttgttgttg	tttgttttgg	gttgtgtttt	aggtgtggag	5460
tttttttgtt	gtggggagag	ttaggggatg	taatttttgt	tgagttttta	agtttaagttg	5520
tttttgtttt	ttttggaagg	tttaagttaa	aaagtttggg	gatggaaagt	tagtgggtaa	5580
atgaagatat	gggatgtggg	tagaagggtg	ttatttagag	tgtttttagg	gagtagggtt	5640
ttaagttttta	aagtgaataa	agagtgggtg	aagatttttt	tttttttttt	tttttttttt	5700
taagaattttt	tttaataagg	aaagttaatg	ttgatttgtt	tttgttttgt	ttttttttat	5760
gtggtagtgt	tgatagagaa	gtgttaagag	tgatagggtg	aggttaggtg	tatttagattt	5820
tttgtgttgt	tagtagttgt	tgtagtattg	atgtgttttt	ttgagtgtat	tttttgtaat	5880
gtgtatatgt	atattttttg	ggtgggttaa	taggagtttg	gttttgttgt	agtttagttt	5940
taggtatttta	ggtgagtgat	ggattagatt	tgtggttttg	tgtttttttg	ttggtttaatt	6000
attttaaaaat	tagagggtggg	ttttttgtgt	ttgagatgtt	attttgttgt	ggtttttttt	6060
agttttttttt	gtttttgttt	tttttttagt	tttttttttg	gtgtgattga	tgtggttttg	6120
tattaatttag	gatgttttga	gttgtgttgt	agggatttgt	ttgtttgtat	ttatttagtag	6180
tgtgggggttg	ggttatttgt	ttgttgttgt	tattgggttt	atataggtaa	gtttttggga	6240
attttagtttt	tgttttagtt	aaggtgattt	ggttttttag	atgaatttaa	aggtgaagag	6300
atgaggttag	gagttgaagg	tttggggaga	gagagtggaa	tggtttaagaa	gagaaaggta	6360
taaggtattaa	taagatatatt	attttttgtg	ttttatttta	tttattttta	ttttttttat	6420
ttatataaaaa	aggagatatg	ttatttaaaa	ttagaaaatt	tgaaaaatag	taataaatta	6480
tttttttgat	tttaaaattt	ttaaatagtt	tgttaagtga	atgttgttgt	aatttgaaga	6540
agttttaatt	gtaaagaaga	tagagttttt	aaaaggtagg	tttaataaatt	agaaatttag	6600
aagtaaatgg	atttgttaaa	agaaaaattt	tttgatttta	aatgaataat	tgtttgtgtg	6660
tttatttttg	atttatataa	gaataaaaaag	ttgttttaga	ttatgttttt	tgtgtgtttt	6720
attagttttt	agatagaaaa	tatatataat	aagagaaatt	tttaatttagt	gttttttaaaa	6780
tgttgaaagt	ttattttatt	tatttaaatgt	tgattaaagt	atatattttta	gatttttttaa	6840
attttttgtta	tattgtatta	agttttgttt	aatttgagag	agttatgttt	taaatattgat	6900
tttttttgttt	atttttattat	taatttagatt	taaattttata	aagtttgttag	aaataaatat	6960
tttgagtttaa	ttatatatga	aatatgtttt	aatgaatttt	tatatataat	agaatgttgt	7020
taaataatta	atttttaagga	taattttttta	tagttatttt	tttttttttag	tgagtttaag	7080
gttgtttttga	gttattaaag	tttaagttag	tagaaggggt	gtgtgtgagt	taaggggtgaa	7140
aagtttagaa	ttgtgtttta	ttagtataag	taaaattttta	tttatataaa	ataaaaaaaa	7200
ttattttttg	agatatatta	tttttatagt	attgtttttta	agtaaaattta	attttttaaag	7260
aaatttaaaga	aagaaattta	aatatatatta	aaataatttt	tgaaagtttt	tttgtttttt	7320
agtatagggt	agtgggagag	gataaattaa	tttttttttg	gtttttgtat	gggtgatttgt	7380
tttatttttg	agttagtgtt	attatttttt	aatgtgtatt	tgtttgatatt	ttagtttaatt	7440
gattttgtaat	gttagtatga	agtattttta	aaatattttt	ttttttgttt	tgtttataag	7500
attgggaaat	ttatttgatg	tggataaaag	tggatgaagt	agattataaa	tatatattgta	7560
atttatgtgt	tttttttttg	ttttgattat	ttttaaattt	tatttgtaat	ttttttttat	7620
tttaaatttg	tagtttaaaag	atgtatatga	gaattgtttt	ttagtttttt	tttatttagta	7680
ttattttatt	tttaagaataa	tttagttgta	agggagggaat	ttttttatag	taagtttaaa	7740
attagtattt	ttgttttttaa	ttttttattt	tattttattt	tattttatat	atatagatat	7800
ttgttttagag	taaaatatat	ttttatgtga	taggttttgta	ttagttgagg	tttatatatt	7860
tagtttatatt	aggtttttgta	atttttattt	taaattatat	atattatat	agtagtttgt	7920
tggtaaagaa	ggttaaatta	atttatatatt	tgtttattat	ttgggtgttta	aatgatgtat	7980
tttatttttg	agattttgtg	gagaattttt	tttttagatt	ttatagtgtt	ttatgaaga	8040
taatgttttt	atatttttag	tgggtttttta	tttgataaga	ttttaatttg	tttaagtttt	8100
ttaaataagg	gttttaaatg	tttttagttg	tttttttatt	gaattttttt	taattttttt	8160
aagattataa	agtatatgtg	taaagtaaat	attttttttt	attgtattgt	tagttgatga	8220

tttataatta	agttaataag	aatttagttt	ttttttgttg	aatgtgttta	tttaattatat	8280
tttagttttt	ttttttaaat	tttagaatag	ttgtggtttt	tataatatta	tgtttttttaa	8340
agtttttatt	tatgaaggga	ttttattata	ttaaagaatg	aaaaaaattt	ttattgtagt	8400
tagtatatat	agttttttat	tttttgtttt	ttaagattta	aatttttagag	ttgtaaatat	8460
ttttggaagt	ttgggtgtta	atgtttttatt	ttagaaagtt	gagaagtttt	atagagttat	8520
atagatttttt	aaattttattt	tttataaaatt	tatagaatttt	tgataaaaagt	tttgggtgggt	8580
ttatttttatt	gatggaatttt	ttattatgat	aaatatatat	gtatgaagga	ttttaatttag	8640
tttttaaagt	ggttgaaaaa	tttaagggtta	tggtgattgtt	ttttatagtg	ttaatgtgtg	8700
tgagatgttg	gaagtattgg	ggattagtag	tagtttagat	gtttaaaaag	ataagggtgtt	8760
ttaatttgtg	tggattttatt	gaagttaaagt	ggtgaataaaa	gataattatt	tagataatttt	8820
agatttaaagt	aaaagtaaaaa	ttatatatttat	ttgtatatata	atattttatat	ttatttttata	8880
ttatagatat	atatatgtat	atatatatgtg	gttttgtaaa	taattgatttt	aaagtggagg	8940
ttttttttgt	attttttttag	taggagttttt	aatatattttt	taattttttta	attatttttat	9000
a						9001

<210> 5

<211> 9001

<212> DNA

<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 5

tgtaaagtga	ttaggagatt	aggagaatgt	tgaaatttttt	gttggaaaaa	tgtaaagaaa	60
attttttatt	tgagttagtt	gtttatagag	ttagtgtgtg	tgtgtgtgtg	tgtgttttga	120
atataaaatg	gatgtgaata	tatatatata	aatagatatg	gtttttgtttt	tatttttaatt	180
tgaattatttt	agataaattgt	ttttatttat	tatttgattt	taatgggttt	atataaatta	240
ggatattttta	tttttttagg	tatttagggt	gttggttgatt	tttagtggtt	ttaatattttt	300
gtatatgttg	gtattatgag	gagtagttat	gtgtttttgg	gttttttaat	tattttggag	360
gttgattgag	gtttttttata	tatgtatatt	tgttgtgatg	aaagtttttat	tggttagagtg	420
gagttatttag	agttttttatt	aaaatttttgt	gggttttatga	gagatgggtt	tagaaaattta	480
tatgggttttg	tgggggttttt	tgggtttttta	aaataaggga	ttaatatttta	agttttttaaa	540
aatattttgta	gtttttgggt	ttgaatttttg	aaaaataagg	agtgagggtt	tgtgtatatt	600
aattatagtg	gagattttttt	ttattttttta	atgtgatgga	gttttttttat	gaaatgaagt	660
tttaaggggt	atgggtattgt	ggggattata	gttatttttga	ggtttaaaaag	aagaaattgg	720
aatatgatta	gtaaatatat	ttagtagaaa	agagttggat	ttttattgat	ttagttatag	780
gttatgggtt	ggtagtgtaa	tgggaggaaa	tattttatttt	atatatatat	tttatgattt	840
tgggggaatt	agaggaaatt	taataagaaa	atgggttagaa	atattttaaaa	tttttatttta	900
aaagattttta	gtaaatttaga	gtttttattag	attaaaaaatt	attataaatg	taagagtatt	960
gttttttagtg	aaatgttgtg	gggttttgaga	aggagattttt	ttgttaaaatt	tttgggataa	1020
aatgtgtttat	ttaaagtatta	gataatgagt	agaatgtaaa	ttaattttaatt	ttttttttatt	1080
aataggtttgt	tagtgtaatg	tgtataaatt	agtataaaga	ttgtaggatt	taatatagtt	1140
ggatgtatga	gtttttagtta	atgtagatttt	gttatatgag	gatgtgtttt	atttttgagta	1200
ggtgttttgta	tgtgtggaat	ggggtaaaagt	ggaataaaaag	gttaaaaagta	gaaatgttga	1260
tttaaagttt	attatgaaga	aattttttttt	ttgtagttaa	attatttttta	aagtgggatg	1320
atattgggtga	agaaagatttg	aaaaataaatt	tttatgtgtg	tttttggatt	gtaagttttta	1380
aatggggagg	agttgttagat	agggttttggg	gggtggttagg	gtaaaggaga	gatataaag	1440
ttgtaaatat	atttgttagtt	tgtttttattt	atttttgtttt	atattgaata	agtttttttaa	1500
ttttgtgaat	aaggataaagg	agggagtggt	ttaaagatat	tttatgttgg	tattgtaaat	1560
tattgatttgt	aatgttaaat	aaatatatat	ttagagatga	taatatattat	tttatagtaa	1620
aataatttgtt	tatgtagaaa	tttagaggag	attagtttgt	ttttttttagt	tgattttatgt	1680
tggggggataa	aaggatttttt	aaaaattattt	ttgaatatgt	ttggatttttt	tttttttaatt	1740
ttttttggaaa	ttaaatttgt	ttggaaatag	tgtttataaag	agttgatgtt	tttaaagggtg	1800
atttttttttg	ttttatataa	ataaggttttt	gttttttgta	gttgagtgtta	gtttttagggt	1860
ttttgtttttt	agttttatata	tattttttttt	gtttgttttg	atttttaatgg	tttaagatag	1920
ttttgagtttt	attgggaaaa	gaaaatgatt	gttaaaaaatt	attttttgaaa	ttgggtatttt	1980
ggtaatatatt	tttaattgtat	ggaaattttat	taagggtatat	tttatatata	attagttttta	2040
ggttggttgat	tttatagggtt	ttatggatttt	aaattttgatt	gataataaag	taaataagag	2100
agttgaatttt	aaagtgtggt	tttttttgggt	taggatgagt	ttaatatagt	gtataaggaa	2160
tttgaaagat	ttaggatattg	tgtttttaatt	aatgtttaagt	agaatggata	agtttttagt	2220
atttttgaaaa	tgttgggttta	gggtttttttt	tttatttgtgt	gtttttttgtt	tggggatttaa	2280
taagtatttat	agagaatgtg	attttgagggtg	atttttttatt	tttgtataaaa	tttagagtga	2340
attatttaaat	agttgttttgt	ttaaagttaa	ggtaatttttt	ttttgatggg	tttattttggt	2400
tttttgatttt	taattttatta	gtttgtttttt	ttagggttttt	gtttttttttt	taattaaagt	2460
tttttttagat	tagtgtagta	tttattttgat	agggtgttttg	gaaaattttta	gattggagag	2520
gtgattttgtt	gttggtttttt	aaattttttta	gtttttaagta	atgtgtttttt	tttttatatg	2580
gggtgggggga	ttggaaatgg	atgtagtgag	atataaagag	tgggtgttttt	gttgatttttt	2640
gtatttttttt	tttttttgatt	atttttatttt	tttttttttaa	gttttttgatt	tttagttttta	2700
tttttttttatt	tttgggttttg	tatttaaaagt	tggattgtttt	tgggttgggtt	aggagttgaa	2760

tttttgggag	tttgtttgtg	tagatttagt	gtgtatgggt	aggtagtagt	ttggtttttgt	2820
attgtttgata	ggtgtaggta	ggatagtttt	tttatttgtg	tttggggtgt	tttgattgggt	2880
gtggagttat	gttagttgta	tttggagaag	ggttttgggag	gaggtggagg	tggagaggggt	2940
tggggagggt	tgtggtggag	tgatgttttt	gtattaggaa	gtttgttttt	ggttttaaga	3000
tgttaggtta	atagggaagt	gtggagttgt	agatttgggt	tgttgtttgt	ttgggtgttt	3060
ggagttgagt	tgtggttaagg	tttggttttt	gtttgattgt	ttgaggggtg	tgtgtgtgtg	3120
tgttgtggag	ggtgtgttta	gaggggtgtg	ttgtggttgt	agtgggttgt	gttgtttag	3180
gggatttaaat	attattttatt	tgtttttgtt	attttttgata	tttttttgtt	aggggtgttg	3240
tgtggggggg	gggtgggtag	agtgtgggtg	gtgttagttt	tttttatttg	aggggttttt	3300
gggggaggga	gggagagaag	aaggggggtt	ttgtttattt	ttgttttgtt	ttggagtgtg	3360
gaagtttgtt	ttttaaaagt	gttttgagt	gtgttttttt	gtttatattt	tatgtttttg	3420
tttgtttgtt	gattttttgt	ttttggattt	ttttgtttga	gttttttggg	ggagatgggg	3480
gtagtttgggt	ttgagaattt	ggtgggggtt	gtgttttttt	gttttttttg	tagtggggaa	3540
attttgtgtt	tagagtgtga	tttggagttg	gtagtgggtg	ttatgggggt	ttggtgggtg	3600
agtagttaag	gatttagtaga	gtgttgtgtt	tttttgttta	tgaattgtat	gaaaggtttg	3660
ttttattttgg	agtattgagt	agtggggatt	aagtgtgttg	ttgttttttt	atttttttgt	3720
tattatttttt	agttgttagt	tatggttttg	gttttgggtt	ttgggttagt	ttgggtgttg	3780
gatttttttt	agtataggtt	ggaggtgtat	attatttttt	atatttttag	tttggagggt	3840
gtaggtaagg	tgttgtgttg	ttttgtagat	attttttgtt	agttgttttg	gtttatttgt	3900
ttttttttgt	tttaagggaag	ttagtttttt	tggggggagg	tgtggtggga	gtgggtgttt	3960
gtttgggttt	ttgtagaatt	tttgggagtt	ggaattttga	ttatttttga	tttttttagt	4020
ttttttttga	ttgggtttgt	ttttgggggt	tttaagggtg	gagtaatttt	gttgtttttt	4080
ttatttgtat	tttgggtttt	tttttgttta	ttgggttata	aaaattttag	tattttgatt	4140
tgaggatttt	tagaggttgt	tgattttttg	ttttgttttt	tttttgggtt	ttagtttttg	4200
aggagtttta	tttgttagga	aattgtttga	aatatttttag	aaatgttttt	tgtgaagagg	4260
tatttttttt	ttttttttgg	gaaagggttg	gtgaattttg	gtgtttaatt	gaatttttat	4320
attttttttt	agtttttttt	aattgtatgg	aaatttgagt	tttttgtgag	ggggaggggg	4380
gtttgtaaat	tatgtgtgtg	tgtgtgtttt	aggagatttg	gtgtgtttgt	gtagaggtgt	4440
ataaatatat	ttgaaagtat	aggttataaa	agtgaatgtg	ttgttgtagt	gagataaata	4500
tgtaaaataa	atgtgtgtgt	ttggggggagg	ggaggaaatg	gggtgtggat	atttatattt	4560
gtgtttgtat	attttatagg	tgtagtgttt	tttgtgggtt	ggagtgtgtg	tgtgtatttt	4620
tttttgggtg	taggtagttt	agttttttta	tggtttttgt	tgttgggtta	gttgggtgtt	4680
gtgttgttag	tgggtatgtt	gatgggaaag	tgtgtgtgtt	ttgtttttag	agaaagataa	4740
aagttagtag	gggaagaatg	aggatgtggg	tgttgaggat	ttgtttaaga	agaagtggta	4800
aagggtggtag	tggattttatt	ttattagtta	gtagtttttag	gagtgtggagg	ttatttttta	4860
gaggaattgt	tattttggata	tgtttatatg	tgaagaaatt	gttgtgtgga	ttatttttat	4920
ggaagtttga	gtttgggtag	gagttagtat	ggagttttgg	agggatgggg	ggaggatgtt	4980
gtggagggtat	aggttaaagta	gattaggaga	gaatgtggaa	ggtagtgttg	tttggggagg	5040
tgttgggtggg	gtgtagttttt	gtaaaggtag	aaggttttgt	ggtgggtttg	ttgtgagatt	5100
atagtttttt	ttttgagggt	gataggattt	ttgtttttgt	ttagggtttt	agagtgggtt	5160
tggttttatt	ttttgttatt	ttgtgatttt	atgagttggg	ttgtatgggt	aattttttgt	5220
ataggatatt	gtgttttttg	tttgtagtgt	ttagagtaga	gttaataaaa	tttttattag	5280
gttaagaggt	gtgaataggt	tttaattttg	gagtttttta	taaggaaaaa	ttgttagaga	5340
tatggaagag	ttgggtttttt	ttgggaaatt	tttgtttttg	ttttgggtta	gttttttttt	5400
ttttgggttt	gtgttttttt	tatttttttt	atgggttgtt	tgggttattt	gggttttttt	5460
atataatttt	tttttttagt	ttgtgatttt	tgggagtaaa	gttttaatat	ataattatta	5520
gttttttttag	aaggagaaag	aaaaaaaaga	gaaagatttt	tttgtttgggt	ttattttatt	5580
tttttttagga	gttggaatttt	ggaaattgaa	attttatatt	tttttttttaa	attataatta	5640
tagtttttga	aaaagggttt	attttaattt	tgtagtaaat	ttgtattttt	tggattggta	5700
aaaatgagtt	taaaataaata	attttaattt	aatgttttgg	tttatgttgg	tttgggtgaag	5760
attttaaaatt	tgttaggatt	ttggaagtag	aaaatagaat	taagtaaaat	aagtgggtatt	5820
tagagggtttt	gttggttaaaa	aaaaaaaatt	aagtgttttg	ggtagaaaaa	ataaagtttt	5880
tggtttagagt	agagtaaaata	aaaagaagaa	aataatgata	aaaagaataa	agattaaaaa	5940
gttttttttaa	attagaggga	atgaagatat	tttttgggtg	gtattttgtg	agggtagag	6000
gttatgtttgg	tggataaaaag	gttgggaaga	agttgaaaaa	ggttttagtt	taattgttta	6060
gagtttagagt	tgggtttttg	gtggtgtggg	tttgagtaag	gttagttttt	tattagtttt	6120
tttgtatatt	aagggaatgg	gttttttatg	tatttttttt	gtttgagtaa	agtttagatg	6180
gtttagggtta	gaaatggtaa	gtaattaaag	atagagttta	tgggtttttt	gggatttttt	6240
gaaaatgttt	ttttattttg	tttgtttatt	tgtagtttta	ttttagtgtt	ttgtagtgtt	6300
ggtgttgggt	ttttttttgta	gttgtttttt	tttttagggg	ggttgtttgt	tgagttaagt	6360
gggagtggag	tgtgtttttt	atagtagttg	ggtgtaaaga	ggaaggggga	taaaaaggaa	6420
attaagaatg	aaaggaaaaa	gagaaaaagt	ggattatatg	gttgggtttg	gtggagatgt	6480
gtaatgtgaa	atattattgg	tgttagtttg	gatatttttag	gttaggtttt	tttttaatat	6540
ataaaaagttg	ttgttttggg	tgatagggag	gtttgatgtg	gattgggatt	ggggttgttg	6600
ttgggttatt	ggatatgggt	ggaagtgtgt	tgggtttggg	ggttgtttgt	aaagttaaat	6660
gatttgggtg	ggttttgggt	gtggataggt	ttgtgggtgg	tttagggtaa	agaagaggta	6720
gagtgaagaa	agggggaatt	tttaaaaatta	tttttttttg	gttttttgag	tttaatatgt	6780
taagtttttg	gggttaatga	gttgatgaag	aggtgggttt	ttgtttttta	ttttaggtgt	6840
ttgttaggtg	agaaagagtg	ttggtgtgtt	agtttttgtt	aagggagtat	gtattggggg	6900
gtgggggatg	atagtggagg	ttagggaagg	aaggggaggaa	ttgtgtggga	gaaagagtga	6960
tttttttagtg	tttttttagt	tttttttttt	tatttgtggg	tttgtgggtt	tggaaatggaa	7020

gtaagtttgt	aaggtgtttt	gggaagggtt	ggaaaagttt	gttgttttgt	gtttgtttta	7080
tattaaagtg	ttttggattt	ggagaaatgt	ttggttgagt	gattaaattg	tttgtaggtt	7140
tttatgtgtt	tgggtgaggt	ttgtggtgta	gttttgagtt	ttagtttgta	ggtttagagta	7200
gattaggttt	tttggttttg	gtggagattt	gggttagtaa	ttgaaagttg	gttttggtat	7260
tttgggtgtg	aggggtggtg	agtgaagtga	ggttaggggtg	tgtgagtgtg	ttagtgtgtg	7320
tgttggggga	aggtgggggt	tgggttttga	tggaaagttt	agtaatttgt	attgtggtat	7380
ttgtttgttt	ttttgtttta	attgttttta	ggtttggttt	aagaattggt	gggttaaagt	7440
gagaaagagg	gagtgttaatt	agtaggttga	gttatgtaag	aatgggtttg	ggttgtagtt	7500
taatgggttt	atgtagtttt	atgatgatat	gtatttaggt	tatttttata	ataattgggt	7560
tgttaagggg	tttataattg	tttttttatt	tattaagagt	tttttttttt	tttaattttat	7620
gaatgttaatt	tttttgttat	tatagagtat	gtttttttta	tttaattttta	ttttgtttat	7680
gagtatgttg	tttagtatgg	tgttttttagt	agtgataggt	gttttggggt	ttagtttttaa	7740
tagtttgaat	aatttgaata	atttgagtag	tttggtgttg	aattttgtgg	tgttgatggt	7800
tgtttgtttt	tatgtgttgt	tgtttttttt	gtatgtttat	agggatatgt	gtaatttgag	7860
tttgggttagt	ttgagattga	aagtaaagta	gtattttagt	tttgggttatg	ttagtgtgta	7920
gaatttggtt	tttaatttga	gtgtttgtta	gtatgtagtg	gattgggtttg	tgtgagttgt	7980
atttatagtg	ttgggatttt	aggattttgt	tggatggggg	aattttgttt	ttgaaagatt	8040
gggaattatg	ttgaaggtt	gtgggtatta	aagaaaggga	gagaaagaga	agttatatag	8100
agaaaaggaa	attattgaat	taaagagaga	gtttttttga	ttttaaaggg	atgttttttag	8160
tgtttgatata	tttttattat	aagtattttt	aatagtgtga	aggatatata	tataaataaa	8220
tgtttgattg	gatattgatat	tttaataatta	ttataagttt	gttatttttt	aagttttagta	8280
ttgttaatat	ttaaatgatt	gaaaggatgt	atataatttg	aaatgttaaa	tttaattttat	8340
aaaagtagtt	gttagtaata	ttataatagt	gttttttaaag	gttaggtttt	aaaataaagt	8400
atgttatata	gaagtgatta	ggattttttg	tttgtagtga	agggagtgtga	tatattaaat	8460
gttatattgt	atgttttttaa	tatattatta	ttattataaa	aaatgtgtga	atattagttt	8520
tagaatagtt	tttttgggtg	atgtaatgat	gtttttgaaa	ttgttatgta	taattttatt	8580
tgtgtataat	attttgtata	atattattgt	tttatttttt	agtaaatatg	aaataaatgt	8640
gttttatttt	atgggagtaa	aatatattgt	atataaattg	gtttggattt	tttttttttt	8700
tttttgttat	taatttggtt	aggatatttt	agttattgtt	ttttaaataa	attagttttt	8760
tttgtttgtt	tagttaaata	tataaggtag	tagtttttat	ttaaatttgg	tagaaataaa	8820
tgatagttat	ttattagaaa	ttaaaaagaa	aaaaaaagg	atttttgggg	gggaaaagg	8880
ttataaaaatt	taattttgtt	tttttaattt	ttttttggtt	taaattttaga	ggattttatt	8940
atgggttagta	aataatatga	aaaagaaaaa	agaagaaaga	aatttagtaa	gtttattagt	9000
t						9001

<210> 6
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PRIMER

<400> 6

gtaggggagg gaagtagatg tt

22

<210> 7
 <211> 24
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PRIMER

<400> 7

ttctaattcct cctttccaca ataa

24

<210> 8
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PROBE

<400> 8

agtcggagtc gggagagcga

20

<210> 9
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> PROBE

<400> 9

agttggagtt gggagagtga aaggaga

27

<210> 10
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> PRIMER CONTROL

<400> 10

tggtgatgga ggaggttag taagt

25

<210> 11
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> PRIMER CONTROL

<400> 11

aaccaataaa acctactcct ccotaa

27

<210> 12
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> PROBE CONTROL

<400> 12

accaccaccc aacacacaat aacaaacaca

30

